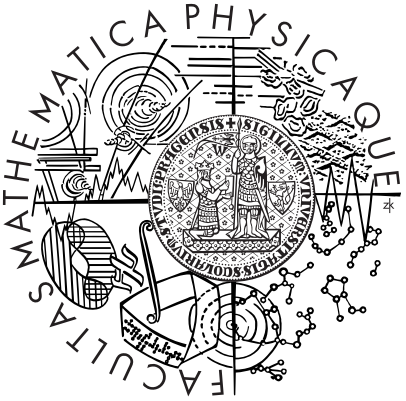
Univerzita Karlova v Praze
Matematicko–fyzikální fakulta

DIPLOMOVÁ PRÁCE



Tomáš Marada

Modelování očekávané ztráty

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Petr Franěk, Ph.D.

Studijní program: Matematika,
Studijní obor: Pravděpodobnost, matematická statistika a ekonometrie

2009

Rád bych poděkoval panu doktoru Petru Fraňkovi, který mě uvedl do oblasti kreditního rizika a motivoval mě k vytvoření této práce. Děkuji za jeho cenné připomínky a kritický praktický pohled. Také bych rád poděkoval paní Heleně Bokové, která mi pomohla s agnličtinou.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 14. duben 2009 Tomáš Marada

Název práce: Modelování očekávané ztráty
Autor: Tomáš Marada
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí diplomové práce: RNDr. Petr Franěk, Ph.D.
e-mail vedoucího: PFranek@csas.cz

Abstrakt: V této práci popisujeme základní principy modelování kreditního rizika. Je zde popsána veškerá nutná matematická teorie. Více se zaměřujeme na modelování Markovských řetězců — především časově homogenní Markovské řetězce se spojitým časem. Hlavní přínos práce je rozšíření modelování Markovských řetězců do náhodného času. Takto pozměněný řetězec nám dává větší svobodu modelování a umožňuje do modelu jednoduchým způsobem vnést a modelovat časovou nehomogenitu. V práci jsme odvozujeme celou řadu odhadů parametrů při různých přístupech k modelování času. Spolehlivost těchto odhadů demonstrujeme na reálných datech a následnou simulací. Ke konci práce popisujeme možné směry dalšího výzkumu.

Klíčová slova: Markovův řetězec, změna času, náhodný čas, Lévy subordinator, kreditní riziko, kraditní ratingy, migrace.

Title: Expected Loss Modelling
Author: Tomáš Marada
Department: Department of Probability and Mathematical Statistics
Supervisor: RNDr. Petr Franěk, Ph.D.
Supervisor's e-mail address: PFranek@csas.cz

Abstract: In this work we describe common credit risk models including all necessary mathematical theory. We extensively study Markov chains, especially homogeneous continuous–time Markov chains. The main contribution of the thesis is an extension of Markov chain modeling into stochastic time — so called time change. This extension allows us to capture better the system dynamics and introduce inhomogeneity into the model in a very elegant way. For practical modeling we derive many parameter estimators under different approaches to the time modeling. Further, we demonstrate the performance of these estimators on real data and simulation. At the end of the thesis we suggest some directions for further research.

Keywords: Markov Chain, Time Change, Stochastic Time, Lévy Subordinator, Credit Risk, Credit Ratings, Migration.

Table of Contents

1 Introduction **1**

2 Stochastic Processes Theory **3**
2.1 Stochastic Processes . . . 3
2.1.1 Stochastic Processes . . . 3
2.1.2 Counting Processes . . . 4
2.1.3 Survival Analysis . . . 5
2.1.4 Lévy Processes . . . 6
2.1.5 Time–inhomogeneous Lévy Processes . . . 9
2.1.6 Affine Processes . . . 10
2.2 Markov Chains Theory . . . 11
2.2.1 Discrete–time Markov Chains . . . 11
2.2.2 Continuous–time Markov Chains . . . 13
2.2.3 Semi–Markov Chains . . . 17
2.3 Partially Observed Data in Markov Chain Theory . . . 18
2.3.1 DA and WA Method . . . 18
2.3.2 Quasi–optimization Method . . . 19
2.3.3 Expectation Maximization Method . . . 20
2.3.4 Markov Chain Monte Carlo Method . . . 22
2.3.5 Componentwise Optimization Method . . . 23

3 Time–changed Continuous–time Markov Chains **25**
3.1 Time–changed Continuous–time Markov Chains . . . 25
3.1.1 Matrix Exponential . . . 26
3.1.2 Continuous–time Markov Chains under Stochastic Time . . . 27
3.2 Maximum Likelihood Estimation . . . 29
3.2.1 Time Evolution Exogenously Given . . . 30
3.2.2 Time Evolution as the Model Parameter . . . 33
3.2.3 Time Evolution as Random Variables . . . 34
3.3 Partially Observed Data under Stochastic Time . . . 38
3.3.1 Quasi–optimization Method . . . 38
3.3.2 Expectation Maximization Method . . . 39
3.3.3 Markov Chain Monte Carlo Method . . . 40
3.3.4 Componentwise Optimization Method . . . 41

4 Credit Risk Models **43**
4.1 Idea of Credit Risk . . . 43
4.2 From Loan to Defaultable Zero–coupon Bond . . . 45
4.2.1 Zero–coupon Bonds . . . 45
4.2.2 Loan as Defaultable Zero–coupon Bond . . . 47
4.3 Structural Models . . . 47
4.3.1 Merton's Model . . . 48
4.3.2 First–Passage Model . . . 50
4.4 Reduced–form Models . . . 54
4.4.1 Default Probability . . . 55
4.4.2 Credit Rating Migration . . . 56
4.4.3 Recovery Rates . . . 59

4.5 Incomplete Information Models . . . 62
4.6 Dependent Default Modeling . . . 63
4.6.1 Unit Loss and Homogeneous Portfolio . . . 64
4.6.2 Factor Models . . . 64
4.6.3 Bernoulli Models . . . 65
4.6.4 Monte Carlo Simulation . . . 69

5 Real Data Study **71**
5.1 Data Analysis and Estimations . . . 71
5.1.1 Fitting Time Evolution by Distribution . . . 74
5.2 EM and CO Method Simulation Comparison . . . 75

6 Summary **79**

Appendix **81**
A Lévy Processes . . . 81
A.1 Lévy Processes . . . 81
A.2 Examples of Lévy Processes . . . 82
A.3 Lévy–Ito Decomposition and Lévy–Khintchin Formula . . . 85
B Copulas . . . 89
C Frequency Data . . . 90

Bibliography **97**

Used Notation

Mathematical Notation

$\mathbb{N}$ – set of natural numbers
$\mathbb{N}_0$ – $\mathbb{N} \cup \{0\}$
$\mathbb{R}$ – set of real numbers
$\mathbb{R}_+$ – positive real numbers
$\mathbb{R}^n$ – n-dimensional real space
$\mathbb{C}$ – set of complex numbers
$\Re(x)$ – real part of x
P – probability measure
P^* – risk–neutral probability measure
$\mathsf{E}\, X$ – expected value of X with respect to measure P
$\mathsf{E}^* X$ – expected value of X with respect to measure P^*
$\mathsf{Var}\,(X)$ – variance of X
$\mathsf{Cov}\,(X, Y)$ – covariance of X and Y
$\mathsf{Corr}\,(X, Y)$ – correlation of X and Y
$\mathcal{L}(X)$ – law of random variable X
$\mu * \nu$ – convolution of μ and ν
$\mathbf{Q}$ – matrix $\mathbf{Q}$ (bold capital notation)
$\boldsymbol{x}$ – vector $\boldsymbol{x}$ (bold slanted notation)
$\|\boldsymbol{x}\|$ – norm of $\boldsymbol{x}$ (Euclidean if not stated otherwise)
$\propto$ – equality up to scaling constant
$\stackrel{d}{=}$ – equality in distribution
$\mathbf{1}_{\{A\}}$ – indicator of event A
$(X)^+$ – positive part of X

Shortcuts

ML – Maximum likelihood
CTMC – continuous–time Markov chain
DA – Diagonal adjustment method
WA – Weighted adjustment method
QO – Quasi–optimization method
CO – Componentwise optimization method
EM – Expectation–maximization method
MCMC – Markov chain Monte Carlo method
TEEG – time evolution exogenously given setting
TEPG – time evolution parametrically given setting
TERV – time evolution as random variables setting
DP – default probability

RR – recovery rates
LGD – loss given default
EAD – exposure at default
RFV – recovery of face value
RT – recovery of treasury
RMV – recovery of market value

1
Introduction

Credit risk has become a big issue in the last few decades, but the need for credit risk measurements is still present since people have started borrowing any good that they need (not just money). In our context the subject of debt is money and the lender is the bank, but the principle is the same in any situation when the subject of debt is anything else than money.

When the bank borrows some amount of money to a debtor, the debtor is obliged to return the full amount plus some compensation for borrowing the money. Of course, there always exists a danger that the debtor will fail to fulfill this obligation. Reasons for failure can be financial distress of the debtor, fraud or something else. In those cases the lender loses some fraction of his money or all of it. How big fraction he loses depends on the debtor's financial situation, willingness of returning money, bankruptcy costs, the obligation of the debtor to other counterparties and on many other circumstances. The lender should be aware of the risk. If he acts rationally, he will lend money only in the case when the profit from lending the money is higher than the expected loss from the provided loan plus some risk premium for bearing the risk. Therefore, he needs to estimate the probability distribution of the credit loss for the entire portfolio of loans.

After estimation of the credit loss distribution the bank would like to hedge against the risk. Here comes into play the rich variety of credit risk derivatives which transfer the credit risk to the third counterparty. Since these credit derivatives provide some protection against the credit risk, the bank is obliged to pay for it. Hence the profit from providing loans is smaller but without a credit risk.

There are two points of view on credit risk models that are closely connected. The first point of view is estimating the expected loss under a real world probability measure that is useful for a bank asset–liability management. The second one is estimating the loss under a risk–neutral probability measure,[1] that is useful for a risk neutral pricing of credit derivatives.

The first approach to credit risk modeling is important for a bank manager who should be interested in the following questions. What is the expected loss from a loan portfolio in next five years? What is the probability of default of a particular debtor? What is the correlation between particular loans? What is the VaR (CVaR)[2] of the loan portfolio? How the bank's loan portfolio depends on the evolution of the economy? We will try to answer some of these questions in the presented work. We will refer to this approach as the portfolio models (the second approach are the pricing models).

A different point of view is adopted by the trader who wants to enter into a contract with the bank and sell a protection against the credit risk. His main issue is what price he should charge for the protection. Of course, the bank, as a protection buyer, is also interested in a fair valuation of the derivative. Many models for the credit derivatives pricing are known. These models can be divided into two groups — structural models and reduced–form models. There also exist some hybrid models that try to integrate both, the structural and the reduced–form approach. Pricing models see the debt as a defaultable zero–coupon bond or as some structure build from it. Hence the main issue is how to price a defaultable zero–coupon bond. Roots of structural models go back to the work of Black and Scholes (1973) and Merton (1974),

[1] For a definition of a risk–neutral measure and conditions for existence see Delbaen and Schachermayer (2006).

[2] Var – Value at Risk, CVaR - Conditional Value at Risk, for a definition see Chapter 4.

later extended by Black and Cox (1976) and by many others. Reduced–form models were introduced mainly by Jarrow and Turnbull (1995), Jarrow et al. (1997), and Lando (1998) and were extended by many others. One of the most widespread approach integrating both approaches together is the incomplete information approach started with Duffie and Lando (2001). A nice short introduction to pricing models is given in Giesecke (2004). All these pricing models try to explain spreads of defaultable zero–coupon bonds. For more details about credit risk derivative pricing we refer to Bielecki and Rutkowski (2002), Duffie and Singleton (2003), Schönbucher (2003), and Lando (2004). The pricing modeling and portfolio modeling can not be separated, but we are going to study mainly portfolio modeling issues and the pricing of credit derivatives will be mentioned just marginally. For portfolio models there is much less available literature. One of the most recent books is for example Bluhm et al. (2002).

There are a lot of problematic issues in the credit risk modeling. First of all, one needs to define a default. For many purposes a different definition is used. One possibility is the random time when a firm's value will drop below some default barrier; then we are dealing with a hitting time theory. A different possibility is the time when the obliger is delayed in fulfilling his obligations for some given amount of time. Even if we have some definition of a default one needs to estimate the exposure at default[3] and the recovery at default.[4] For these estimations banks have very little data. The lack of data is one of the biggest problems in the credit risk modeling.

This work has the following structure. In the second chapter we give a necessary mathematical theory about tools used in the credit risk modeling, especially Markov chains. In the third chapter we extend the Markov chain theory into a stochastic time. We derive some maximum likelihood estimators in three possible approaches to time evolution. The maximum likelihood estimator can be used only in the case of continuously observed data which is often not the case in credit risk. Hence, we study few methods that can be used in the situation of partial data and we extend few of them for stochastic times. In the forth chapter we give an overview on commonly used models in the portfolio management and in the fifth chapter we do a real data study where we show the performance of our method. Introducing a stochastic time evolution into the Markov chain modeling gives us a 35 % better fit to data. We think it is a very nice improvement. At the end of the thesis we give a summary and few possible future directions of our research.

[3]The height of the debt at time of default.

[4]The fraction of the debt which will be covered at default.

2
Stochastic Processes Theory

In this chapter we recall some known results from the theory of stochastic processes which we will need later. Nevertheless, some elementary probability concepts will be used implicitly and can be found in every introduction to probability books such as Feller (1968), Karlin and Taylor (1975), Karlin and Taylor (1981), Shiryaev (1995), Kallenberg (1997), Chung (2001), and many others.

2.1 Stochastic Processes

In this section we will define general stochastic process, Lévy processes and affine processes. We will also mention basic concept of survival analysis.

2.1.1 Stochastic Processes

Let us assume a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathsf{P})$. A *stochastic process* $(X_t, t \in T)$ is a family of random variables with index set $T \subset \mathbb{R}_+ := [0, \infty)$. In the case when $T = \mathbb{N}_0$, we rather talk about a random sequence denoted by $\{X_n\}_{n\in\mathbb{N}_0}$.
An adapted process X_t with $\mathsf{E}\,|X_t| < \infty$, for every $t \in T$, is called a

- *submartingale* if for every $s < t$

$$\mathsf{E}\,[X_t|\mathcal{F}_s] \geq X_s,$$

- *supermartingale* if for every $s < t$

$$\mathsf{E}\,[X_t|\mathcal{F}_s] \leq X_s,$$

- *martingale* if for every $s < t$

$$\mathsf{E}\,[X_t|\mathcal{F}_s] = X_s.$$

Let N_t be a non–negative right–continuous submartingale with $N_0 = 0$. Then from the Doob–Mayer decomposition (see Karatzas and Shreve (1991, Theorem 1.4.10) or Kallenberg (1997, Theorem 22.5)) there exists a unique right–continuous martingale M_t and a right–continuous non–decreasing predictable process A_t such that

$$N_t = M_t + A_t, \qquad t \geq 0,$$

with $A_0 = 0$. The process A_t is called the *compensator* of the process N_t. If A_t is differentiable it can be written in the form

$$A_t = \int_0^t \lambda_s \,\mathrm{d}s,$$

where λ_s is a non–negative right–continuous predictable process. This form will be useful later.

If N_t has independent increments, then A_t is a deterministic function equal to expected value of N_t

$$A_t = \mathsf{E}\,N_t, \qquad t \geq 0. \tag{2.1}$$

2.1.2 Counting Processes

Let $\{T_n\}_{n\in\mathbb{N}}$ be an increasing sequence of random times with values in $(0,\infty)$ and

$$\mathsf{P}\left[\lim_{n\to\infty} T_n = \infty\right] = 1.$$

Then we can define an *associated counting process* N_t as

$$N_t = \sum_{n=1}^{\infty} \mathbf{1}_{\{T_n \le t\}}.$$

The counting process N_t is a process that counts the number of events that have arrived up to time t. It starts at 0 and is non–decreasing with jumps of size 1 (the possibility of two events arriving at the same time is excluded). From that it follows that it can take values only from $\mathbb{N}_0$.

Random times between event arrivals are called *inter–arrival times* and are denoted by

$$\begin{aligned} \tau_1 &= T_1, \\ \tau_i &= T_i - T_{i-1}, \qquad i > 1, \end{aligned}$$

If the inter–arrival times $\tau_1, \tau_2, \ldots$ are independent identically distributed random variables having an exponential distribution with parameter λ, then N_t is called a *Poisson process* with intensity λ. Since the inter–arrival times $\tau_1, \tau_2, \ldots$ are independent exponentially distributed, N_t has independent stationary increments with the Poisson distribution

$$\mathsf{P}[N_t - N_s = k] = \mathsf{P}[N_{t-s} = k] = \frac{\big(\lambda(t-s)\big)^k}{k!}\mathrm{e}^{-\lambda(t-s)}, \qquad s < t, k = 0, 1, \ldots.$$

Since a Poisson process has independent increments and the mean value is $\mathsf{E}\, N_t = \lambda t$, we conclude from (2.1) that the compensator of the process N_t is $A_t = \lambda t$, and process $N_t - \lambda t$ is a martingale. Stochastic process $N_t - \lambda t$ is called a *compensated Poisson process*. As a matter of fact, it is not hard to prove that the only counting process with stationary independent increments is the Poisson process.

In general, the intensity parameter λ does not need to be constant. In the case when the intensity is a time dependent function, we denote it by λ_t and speak about an *inhomogeneous Poisson process*. Increments of an inhomogeneous Poisson process are not stationary anymore and their Poisson distribution is time dependent in following way

$$\mathsf{P}[N_t - N_s = k] = \frac{\left(\int_s^t \lambda_u \,\mathrm{d}u\right)^k}{k!}\mathrm{e}^{-\int_s^t \lambda_u \,\mathrm{d}u}, \qquad s < t, k = 0, 1, \ldots.$$

We can go even further and allow λ_t to be a stochastic non–negative process such that if we condition on a particular $\omega \in \Omega$, the counting process N_t with intensity $\lambda_t(\omega)$ becomes an inhomogeneous Poisson process. This process is called a *Cox process* or a *doubly stochastic Poisson process*. Then, it holds

$$\mathsf{P}[N_t - N_s = k] = \mathsf{E}\left[\frac{\left(\int_s^t \lambda_u \,\mathrm{d}u\right)^k}{k!}\mathrm{e}^{-\int_s^t \lambda_u \,\mathrm{d}u}\right], \qquad s < t, k = 0, 1, \ldots.$$

For more technical details about the Cox process see Grandell (1976) or Kallenberg (1997, Chapter 10).

2.1.3 Survival Analysis

Let us assume a non–negative random variable $\tau : \Omega \to \mathbb{R}$ which has the interpretation of a random time before some particular event will happen. For convenience we can assume that $\mathsf{P}[\tau > 0] = 1$. In context of survival analysis τ is often called a *survival time*.

Let us denote the cumulative distribution function of τ by $F(t)$. Then we can define a *survival function* $S(t)$ as

$$S(t) = 1 - F(t) = \mathsf{P}[\tau > t].$$

The survival function shows what the probability is that the survival time τ will be bigger than t. Furthermore, we assume that there exists a density function $f(t)$ that is the derivative of $F(t)$. Then let us define the *hazard rate* λ_t as

$$\lambda_t = \lim_{h \downarrow 0} \frac{\mathsf{P}[t \leq T < t + h | T > t]}{h} = \lim_{h \downarrow 0} \frac{F(t+h) - F(t)}{h} \cdot \frac{1}{1 - F(t)} = \frac{f(t)}{1 - F(t)}.$$

The hazard rate λ_t shows the conditional intensity with which the random event will occur in an infinitesimal time increment after t, conditioned that the random event has not occurred yet. For λ_t the following identities hold

$$\lambda_t = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{\partial \log S(t)}{\partial t}. \tag{2.2}$$

If we know the hazard rate, we can easily recover the cumulative distribution function and the survival function from (2.2) as

$$S(t) = \exp\left(- \int_0^t \lambda_s \, \mathrm{d}s \right).$$

It is clear that if and only if the hazard rate is constant, the random time T has an exponential distribution with intensity λ.

In survival analysis one often uses a more general Weibull distribution with cumulative distribution function

$$F(t) = 1 - \mathrm{e}^{-(ct)^k}.$$

The exponential distribution is a special type of the Weibull distribution with $k = 1$. The hazard rate for the Weibull distribution is

$$\lambda_t = ck(ct)^{k-1}.$$

Let $N_t = \mathbf{1}_{\{\tau \leq t\}}$ be the counting processes associated with the random time τ. It is obvious, that the counting process N_t is a right–continuous non–negative submartingale with $N_0 = 0$, hence there exists a non–decreasing compensator A_t of the process N_t. Let us assume that λ_t is the hazard rate of the random time τ, then

$$A_t = \int_0^{t \wedge \tau} \lambda_s \, \mathrm{d}s = \int_0^t \lambda_s \mathbf{1}_{\{s \leq \tau\}} \, \mathrm{d}s.$$

The compensator A_t describes a cumulative conditional likelihood of a default. It is obvious that the counting process N_t is a Poisson process (resp. the time inhomogeneous Poisson process, resp. a Cox process) stopped at random time τ if the hazard rate λ_t is constant (resp. deterministic, resp. random). The stochastic dynamics of the counting process is then

$$\mathrm{d}N_t = (1 - N_t)\lambda_t \, \mathrm{d}t + \mathrm{d}M_t,$$

where M_t is a martingale.

2.1.4 Lévy Processes

In this section we define Lévy processes and Lévy subordinators and describe only their most necessary properties. For reader, who is not familiar with the theory of Lévy processes, we recommend to read Appendix A, where the Lévy processes theory is described in a scope relevant to this paper. For deeper theory see classical references for Lévy process theory such as Bertoin (1996) and Sato (1999). An overview of the most common Lévy processes with references to original papers can be found in Schoutens (2003, Chapter 5). Lévy subordinators are described for example in Bertoin (1999).

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathsf{P})$ be the probability space equipped by filtration $\mathcal{F}_t, t \geq 0$ that satisfies the "usual hypothesis", i.e., is right–continuous and for every $t \geq 0$ the σ-algebra $\mathcal{F}_t$ contains all null sets of the σ-algebra $\mathcal{F}$. We will describe properties for a one–dimensional Lévy process. For multi–dimensional properties see the literature above.

A stochastic cadlag[5] process X_t adapted to filtration $\mathcal{F}_t$ with values in $\mathbb{R}$ is a Lévy process if and only if it has following properties:

(L1) $X_0 = 0$ a.s.

(L2) For every sequence of times $t_0 < t_1 < \ldots < t_n$ the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.

(L3) The process X_t has stationary increments, i.e., for every $t > 0$ and $s > 0$

$$\mathcal{L}(X_{t+s} - X_t) = \mathcal{L}(X_s),$$

where $\mathcal{L}(X)$ denotes the law of random variable X.

(L4) The process X_t is continuous in probability, i.e., for every $\varepsilon > 0$ holds

$$\lim_{h\to 0} \mathsf{P}[|X_{t+h} - X_t| > \varepsilon] = 0.$$

In other words, Lévy process is a stochastic process starting at 0 with independent stationary increments which is continuous in probability. In Appendix A is shown that every Lévy process can be decomposed into two parts. First part is a Brownian motion with drift $b \in \mathbb{R}$ and variance $\sigma \geq 0$. The second part is pure jump process with *Lévy measure* ν. For derivation of Lévy measure see Appendix A.

For characteristic function of X_t holds well–known Lévy–Khintchine formula

$$\begin{aligned}\mathsf{E}\,\mathrm{e}^{izX_t} &= \mathsf{E}\,\mathrm{e}^{t\psi(z)} \\ &= \exp\left(t\left(ibz - \frac{1}{2}\sigma^2 z^2 + \int_{\mathbb{R}} \left(\mathrm{e}^{izx} - 1 - izx\mathbf{1}_{\{|x|<1\}}\right)\nu(\,\mathrm{d}x)\right)\right), \qquad z \in \mathbb{R},\end{aligned}$$

where the Lévy measure ν satisfies the following two properties

$$\begin{aligned}\nu(\{0\}) &= 0, \\ \int_{\mathbb{R}} (|x|^2 \wedge 1)\,\nu(\,\mathrm{d}x) &< \infty.\end{aligned}$$

Function $\psi(z)$ is called the *characteristic exponent*. For understanding where this formula came from see Appendix A. If the Lévy measure ν satisfies

$$\int_{|x|<1} |x|\,\nu(\,\mathrm{d}x) < \infty,$$

[5] The cadlag process is a process with right–continuous sample path which admits a left limit.

we can rewrite the characteristic exponent in the Lévy–Khintchin formula in the form

$$\psi(z) = ib'z - \frac{1}{2}\sigma^2 z^2 + \int_{\mathbb{R}} \left(\mathrm{e}^{izx} - 1\right) \nu(\,\mathrm{d}x),$$

where

$$b' = b - \int_{|x|<1} izx\ \nu(\,\mathrm{d}x).$$

Further, note that general Lévy process can be decomposed into Brownian motion with drift and pure jump process. For details see Appendix A.

Lévy Subordinator

A *Lévy subordinator* is a one–dimensional Lévy process which is almost surely increasing and can be interpreted as a random time evolution. It is clear that if a subordinator T_t is almost surely increasing then the Brownian part of the process has to be zero, the drift b has to be non–negative, and the process can have only positive jumps. A subordinator has finite variation and hence the characteristic exponent in the Lévy–Khintchin formula takes the form

$$\psi(z) = ibz + \int_0^\infty \left(\mathrm{e}^{izx} - 1\right) \nu(\,\mathrm{d}x), \qquad z \in \mathbb{R},$$

where

$$\begin{aligned} b &\geq 0, \\ \nu\big((-\infty, 0)\big) &= 0, \\ \int_0^\infty (1 \wedge x)\nu(\,\mathrm{d}x) &< \infty. \end{aligned}$$

In the case when T_t is a subordinator, it is more convenient to work instead of the characteristic exponent ψ with the Laplace exponent η, which is defined as

$$\eta(u) = -\psi(iu) = bu + \int_0^\infty \left(1 - \mathrm{e}^{-ux}\right) \nu(\,\mathrm{d}x), \qquad u \in \mathbb{C}, \Re(u) \geq 0.$$

Then it holds that

$$\mathsf{E}\,\mathrm{e}^{-uT_t} = \mathrm{e}^{-t\eta(u)}, \qquad u \in \mathbb{C}, \Re(u) \geq 0.$$

Sometimes it is useful to work with the Laplace exponent $\eta(u)$ for u such that $\Re(u) < 0$. According to Sato (1999, Theorem 25.17), $\mathsf{E}\,\mathrm{e}^{-uT_t} < \infty$ if and only if

$$\int_1^\infty \mathrm{e}^{-ux}\nu(\,\mathrm{d}x) < \infty.$$

The Laplace exponent is defined on the interval $(\bar{u}, \infty)$ or $[\bar{u}, \infty)$. The left endpoint $\bar{u}$ may or may not belong to the interval on which the Laplace exponent is defined, but every time $\bar{u} \leq 0$.

Note that since a subordinator T_t has only positive jumps $\nu((-\infty, 0)) = 0$, and non–negative drift b, it is clear that $T_t \geq bt$. The line bt is a lower bound for the process T_t. We need to consider if we should allow all values from the interval $[0, \infty)$ for T_t, i.e., not have any

lower bound, or if there is some lower bound which can be equipped by the term bt. In most cases there is not such a bound and we should set the drift term equal to zero $b = 0$.

Some authors mean by the term subordinator a more general process T'_t which is a subordinator T_t killed with rate λ. More precisely, let T_t be a subordinator and τ a random variable with exponential distribution having intensity λ independent of T_t. Then the process

$$T'_t = \begin{cases} T_t, & t < \tau, \\ \infty, & t \geq \tau \end{cases}$$

is a subordinator killed with rate λ. As a special case when $\lambda = 0$, we get a classical subordinator T_t. We will not need killed subordinators in the presented work but we mention it just to be clear what we mean by the term subordinator. For more about (killed) subordinators we refer to Bertoin (1999).

Examples of Subordinators

Compound Poisson process One of the simplest subordinator is a Poisson process or a *compound Poisson process*, which is a jump process with exponentially distributed inter–arrival times, like Poisson process, but random jump size. Poisson process has jump size of 1 and hence it is a degenerate case of a compound Poisson process. A compound Poisson process is a subordinator if the probability distribution of jumps is non–negative. If λ is an intensity of a Poisson process and F is the non–negative probability distribution of jumps then the Lévy measure of subordinator T_t is

$$\nu(\mathrm{d}x) = \lambda F(\mathrm{d}x).$$

α-stable subordinator An α-stable subordinator T_t is a subordinator with the Laplace exponent $\eta(u) = u^\alpha$, where $0 < \alpha \leq 1$. If we want to get a Lévy measure of the subordinator we need to notice that

$$\begin{aligned} u^\alpha &= \frac{u^\alpha}{\Gamma(1-\alpha)} \int_0^\infty \mathrm{e}^{-s} s^{-\alpha} \,\mathrm{d}s \\ &= \frac{\alpha}{\Gamma(1-\alpha)} \int_0^\infty u\mathrm{e}^{-uy} \frac{1}{\alpha} y^{-\alpha} \,\mathrm{d}y \qquad (s = uy) \\ &= \frac{\alpha}{\Gamma(1-\alpha)} \int_0^\infty u\mathrm{e}^{-uy} \int_y^\infty x^{-\alpha-1} \,\mathrm{d}x \,\mathrm{d}y \\ &= \frac{\alpha}{\Gamma(1-\alpha)} \int_0^\infty \int_0^x u\mathrm{e}^{-uy} \,\mathrm{d}y x^{-\alpha-1} \,\mathrm{d}x \\ &= \frac{\alpha}{\Gamma(1-\alpha)} \int_0^\infty (1 - \mathrm{e}^{-ux}) x^{-\alpha-1} \,\mathrm{d}x. \end{aligned}$$

We see that the Lévy measure ν is

$$\nu(\mathrm{d}x) = \frac{\alpha x^{-\alpha-1}}{\Gamma(1-\alpha)} \,\mathrm{d}x.$$

The case when $\alpha = 1$ is degenerate, and in that case we get deterministic $T_t = t$.[6]

[6] Note that an α-stable subordinator is a special case of an α-stable process for choice of parameters $\beta = 1$, $b = 0$, and $a^\alpha = \cos(\alpha\pi/2)$ (see Appendix A).

Gamma subordinator Let T_t be a Lévy process with increments that are gamma distributed. We say T_t is a *Gamma process* with density function

$$f_{T_t}(x) = \frac{a^{bt}}{\Gamma(bt)} e^{-ax} x^{bt-1}, \qquad x > 0,$$

where $a, b > 0$. It follows that

$$\begin{aligned}
\mathsf{E}\, \mathrm{e}^{-uT_t} &= \int_0^\infty \mathrm{e}^{-ux} \frac{a^{bt}}{\Gamma(bt)} e^{-ax} x^{bt-1}\, \mathrm{d}x \\
&= \frac{a^{bt}}{(a+u)^{bt}} \int_0^\infty \frac{(a+u)^{bt}}{\Gamma(bt)} e^{-(a+u)x} x^{bt-1}\, \mathrm{d}x \\
&= \left(\frac{a}{a+u}\right)^{bt} \\
&= \exp\left(bt \log\left(\frac{a}{a+u}\right)\right) \\
&= \exp\left(bt \int_0^\infty \frac{\mathrm{e}^{-(a+u)x} - \mathrm{e}^{-ax}}{x}\, \mathrm{d}x\right) \\
&= \exp\left(-t \int_0^\infty b(1 - \mathrm{e}^{-ux}) x^{-1} \mathrm{e}^{-ax}\, \mathrm{d}x\right),
\end{aligned}$$

where we used Frullani's integral.[7] We see that the Lévy measure is

$$\nu(\mathrm{d}x) = bx^{-1}\mathrm{e}^{-ax}\, \mathrm{d}x.$$

2.1.5 Time–inhomogeneous Lévy Processes

Lévy processes have stationary independent increments, which are very useful for computations. We can extend the class of Lévy processes further to the time–inhomogeneous Lévy processes setting, where most of nice properties of Lévy processes remains, but we get a more general class of stochastic processes which still have independent increments, but increments are no longer stationary. Time–inhomogeneous Lévy processes are also called processes with independent increments and absolutely continuous characteristics. For more about processes with independent increments and their characteristics see Jacod and Shiryaev (2003, Chapter 2).

Let T^* be some finite time horizon. A stochastic cadlag process X_t adapted to filtration $\mathcal{F}_t$ with values in $\mathbb{R}$ is an inhomogeneous Lévy process if and only if it has following properties:

(L1) $X_0 = 0$ a.s.

(L2) For every sequence of times $t_0 < t_1 < \ldots < t_n$ the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.

(L3) For all $0 \le t \le T^*$, X_t has a time–dependent characteristic exponent

$$\psi_t(z) = \int_0^t \left(ib_s z - \frac{1}{2}\sigma_s^2 z^2 + \int_{\mathbb{R}} \left(\mathrm{e}^{izx} - 1 - iux\mathbf{1}_{\{|x|<1\}}\right) \nu_s(\mathrm{d}x) \right) \mathrm{d}s, \qquad z \in \mathbb{R},$$

where $b_s \in \mathbb{R}$, $\sigma_s > 0$, and ν_s is a measure on $\mathbb{R}$ that integrate $(|x|^2 \wedge 1)$ satisfying $\nu\{0\} = 0$. Furthermore, we assume that

$$\int_0^{T^*} \left(|b_s| + \sigma_s^2 + \int_{\mathbb{R}} \left(|x|^2 \wedge 1\right) \nu_s(\mathrm{d}x) \right) \mathrm{d}s < \infty.$$

[7] $\int_0^\infty \frac{f(ax) - f(bx)}{x}\, \mathrm{d}x = \big(f(0) - f(\infty)\big) \log \frac{b}{a}$.

The random variable X_t has an infinitely divisible distribution with the characteristic triplet (b_t, σ_t, ν_t). If the characteristic triplet depends on time linearly, i.e., $(b_t, A_t, \nu_t) = (tb, tA, t\nu)$, the process X_t is a time–homogeneous Lévy process.

From the definition we see that for the characteristic function $\phi_{X_t}(z)$ it holds that

$$\phi_{X_t}(z) = \mathrm{e}^{\psi_t(z)},$$

and only in the case of a time–homogeneous Lévy processes we have

$$\phi_{X_t}(z) = \mathrm{e}^{\psi_t(z)} = \mathrm{e}^{t\psi(z)}.$$

Similarly, in the case of a Lévy time–inhomogeneous subordinator, the Laplace exponent η_t is now time–dependent.

Time–inhomogeneous Lévy processes are not very commonly used processes. Some examples can be found in Cont and Tankov (2003, Chapter 14). Sometimes the time interval $(0, T^*)$ is divided into subpartitions on which the process X_t is modeled via a time–homogeneous Lévy processes with different parameters on each subpartition. On the whole interval $(0, T^*)$ the process X_t is a time–inhomogeneous process. More about time–inhomogeneous Lévy processes can be found for example in Kluge (2005).

2.1.6 Affine Processes

A stochastic process X_t is Markov if and only if for every $0 \leq s < t$

$$\mathsf{P}[X_t \leq x | \mathcal{F}_s] = \mathsf{P}[X_t \leq x | \sigma(X_s)], \tag{2.3}$$

where $\mathcal{F}_s$ is a σ-algebra generated by the process X_t up to time s and $\sigma(X_s)$ is a σ-algebra generated by X_s. If the process X_t satisfies (2.3) we say that X_t satisfies the *Markov property*. The Markov property does not say anything different than that the conditional probability distribution of the process X_t conditioned on the whole past till time s is the same as the conditional probability distribution of the process X_t conditioned by knowledge of the value of the process at time s.

The Markov process X_t with some state space $\mathbb{R}_+^n \times \mathbb{R}^{d-n}$, $0 \leq n \leq d$, is called the *regular affine process* if and only if its conditional characteristic function is "exponential–affine", i.e.,

$$\mathsf{E}\left[\mathrm{e}^{iu \cdot X_t} | X_s\right] = \mathrm{e}^{\alpha(t-s,iu) + \beta(t-s,iu) \cdot X_s}, \qquad u \in \mathbb{R}^d,$$

for some functions $\alpha(t, iu)$ and $\beta(t, iu)$ which are differentiable in t and their derivatives are continuous at 0.

Let us assume $d = 1$ and the Markov process X_t follows dynamics

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t,$$

where W_t is a standard Brownian motion and the functions $\mu(x)$ and $\sigma^2(x)$[8] are both affine in x.[9] Furthermore, let $\lambda(x)$ be also affine in x, then there exist functions $A(t)$ and $B(t)$ such that

$$\mathsf{E}\left[\exp\left(-\int_0^t \lambda(X_s)\,\mathrm{d}s\right)\right] = \exp(A(t) - B(t)X_0).$$

[8] Notice that we want $\sigma^2(x)$ to be affine in x not $\sigma(x)$.

[9] We say that function $f : \mathbb{R}^d \to \mathbb{R}$ is affine if there exist $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$, such that $f(x) = a \cdot x + b$.

If we assume $\lambda(x) = x$, $\mu(x) = \kappa(\theta - x)$ and $\sigma(x) = \sigma$ then we get the model of Vašíček (1977)

$$\mathrm{d}X_t = \kappa(\theta - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t,$$

and

$$B(t) = \frac{1}{\kappa}\left(1 - \mathrm{e}^{-\kappa t}\right),$$
$$A(t) = \frac{\left(B(t) - t\right)\left(\kappa^2\theta - \frac{\sigma^2}{2}\right)}{\kappa^2} - \frac{\sigma^2 B(t)^2}{4\kappa}.$$

If we replace in the Vašíček's model function $\sigma(x) = \sigma$ by $\sigma(x) = \sigma\sqrt{x}$, we get the CIR model of Cox et al. (1985)

$$\mathrm{d}X_t = \kappa(\theta - X_t)\,\mathrm{d}t + \sigma\sqrt{X_t}\,\mathrm{d}W_t,$$

and

$$B(t) = \frac{2(\mathrm{e}^{\gamma t} - 1)}{(\gamma - \kappa)(\mathrm{e}^{\gamma t} - 1) + 2\gamma},$$
$$A(t) = \frac{2\kappa\theta}{\sigma^2}\log\left(\frac{2\gamma(\mathrm{e}^{(\gamma-\kappa)t} - 1)}{(\gamma - \kappa)(\mathrm{e}^{\gamma t} - 1) + 2\gamma}\right),$$

where $\gamma = \sqrt{\kappa^2 + 2\sigma^2}$.

More mathematical theory about affine processes can be found in Duffie et al. (2003) for time homogeneous Markov processes and in Filipovič (2005) extended to time inhomogeneous Markov processes.

2.2 Markov Chains Theory

The special case of a Markov process (defined in previous section) is a *Markov chain* which is a Markov process with a countably large state space E. Markov chains are very useful processes for modelling dynamics of every system which can take countably many states. Vast amount of results are well-known. We recall basic definitions and theory which we will need later. For a deeper Markov chain theory see Karlin and Taylor (1975), Karlin and Taylor (1981), Resnick (1992), Kijima (1997), Stroock (2005), or Prášková and Lachout (2005).

2.2.1 Discrete–time Markov Chains

We call a random sequence $\{X_n\}_{n\in\mathbb{N}_0}$ with values in a countable state space E a *discrete–time Markov chain* if and only if it satisfies the Markov Property, i.e.,

$$\mathsf{P}[X_{n+m} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0] = \mathsf{P}[X_{n+m} = j | X_n = i],$$

for all $n, m \in \mathbb{N}$ and for all $i, j, i_0, \ldots, i_{n-1} \in E$ such that

$$\mathsf{P}[X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0] > 0.$$

Let us denote the transition probabilities

$$\mathsf{P}[X_{n+m} = j | X_n = i] = p_{ij}(n, n+m).$$

If the state space E is finite, we call X_n a finite discrete–time Markov chain. If E is countably infinite it can be labeled by natural numbers. Hence we can work with $E = \mathbb{N}$. In credit risk management we are mainly interested in finite Markov chains, therefore in the following we will assume $E = \{1, 2, \ldots, K\}$.

We call X_n a *homogeneous Markov chain* if the transition probabilities $p_{ij}(n, n+m)$ do not depend on n but only on m, i.e.,

$$\mathsf{P}[X_{n+m} = j | X_n = i] = \mathsf{P}[X_m = j | X_0 = i] = p_{ij}^{(m)}, \tag{2.4}$$

for all $i, j \in E$, $n \in \mathbb{N}_0$ and $m \in \mathbb{N}$. If (2.4) does not hold we speak about an *inhomogeneous Markov chain*. In the following we assume X_n to be homogeneous. The matrix

$$\mathbf{P}^{(m)} = \left(p_{ij}^{(m)}\right)_{i,j=1}^{K},$$

is called the m-step transition probability matrix of the chain. It is easy to verify the so–called Chapman–Kolmogorov equations

$$\mathbf{P}^{(n)}\mathbf{P}^{(m)} = \mathbf{P}^{(n+m)}. \tag{2.5}$$

Equation (2.5) gives us a way to compute the n-step transition probability matrix $\mathbf{P}^{(n)}$ from the one–step transition probability matrix $\mathbf{P}$ as

$$\mathbf{P}^{(n)} = \mathbf{P}^n.$$

We call the probability distribution of X_0, the *initial distribution* of the process and denote it by

$$p_i = \mathsf{P}[X_0 = i].$$

Then it follows that

$$\mathsf{P}[X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i_n] = p_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n}. \tag{2.6}$$

Maximum Likelihood Estimation

We see that the one–step transition probability matrix is a building stone for a discrete–time Markov chain modeling. However, in many practical situations we have observations $x_0, x_1, \ldots, x_m$ of a process X_n and we would like to estimate the transition probability matrix $\mathbf{P}$ from these observations. Let us denote by n_{ij} the frequency of transitions from i to j that occurred in observations $x_0, x_1, \ldots, x_m$ and define

$$n_{i.} = \sum_{j=1}^{K} n_{ij}.$$

Since $\mathbf{P}$ is a probability matrix, its row–sums must be 1 and hence we can write

$$p_{iK} = 1 - \sum_{k=1}^{K-1} p_{ik}. \tag{2.7}$$

Using (2.6) and (2.7) the log–likelihood function is given by

$$l(\mathbf{P}) = \log \mathsf{P}[X_0 = i_0, \ldots, X_m = i_m] = \log\left(p_{i_0} \prod_{i,j} p_{ij}^{n_{ij}}\right)$$
$$= \log(p_{i_0}) + \sum_{i=1}^{K}\left(\sum_{j=1}^{K-1} n_{ij}\log(p_{ij}) + n_{iK}\log\left(1 - \sum_{k=1}^{K-1} p_{ik}\right)\right).$$

After differentiation of $l(\mathbf{P})$ and setting the derivatives to 0, we get the following equations

$$\frac{\partial}{\partial p_{ij}} l(\mathbf{P}) = \frac{n_{ij}}{p_{ij}} - \frac{n_{iK}}{1 - \sum_{k=1}^{K-1} p_{ik}} = 0, \tag{2.8}$$

for $i = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, K-1$. By solving (2.8) together with (2.7) we get the maximum likelihood estimator of the one–step transition probability matrix

$$\hat{\mathbf{P}} = \left(\frac{n_{ij}}{n_{i\cdot}}\right)_{i,j=1}^{K}. \tag{2.9}$$

The estimator of the transition probability matrix $\hat{\mathbf{P}}$ is often called the *cohort estimator* or the estimation by the *cohort method*.

2.2.2 Continuous–time Markov Chains

A *continuous–time Markov chain* can be seen as a generalization of a discrete–time Markov chain. In the discrete–time setting the Markov chain can change the state only in some natural multiples of the time unit. If we replace these discrete times by random exponentially distributed variables we get a continuous–time Markov chain, where a change of the state can happen any time.

Let $T = [0, \infty)$ and let the stochastic process $(X_t, t \in T)$ with values in a countable state space E satisfy the Markov property, i.e., for every $s \geq 0$ and $t \geq 0$,

$$\mathsf{P}[X_{t+s} = i | X_u, 0 \leq u \leq s] = \mathsf{P}[X_{t+s} = i | X_s],$$

then we say that X_t is a continuous–time Markov chain. Let us denote the transition probabilities by

$$\mathsf{P}[X_{t+s} = j | X_s = i] = p_{ij}(s, s+t).$$

Similar as in the case of a discrete–time Markov chain we talk about a homogeneous continuous–time Markov chain if the transition probabilities $p_{ij}(s, s+t)$ do not depend on s, but only on t. In that case we write $p_{ij}(t)$ for short. In the following we assume a finite homogeneous chain ($E = \{1, \ldots, K\}$) if not stated otherwise.

One can see, that we need a whole family of transition probability matrices

$$\mathbf{P}(t) = \left(p_{ij}(t)\right)_{i,j=1}^{K},$$

that depend on time t. It is convenient to define $\mathbf{P}(0) = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Let us denote the probability distribution of X_t at time t by

$$\boldsymbol{p}(t) = \left(p_i(t)\right)_{i=1}^{K},$$

where $p_i(t) = \mathsf{P}[X_t = i]$. Then $\boldsymbol{p}(0)$ is the initial distribution of the process and for every $0 < t_1 < \ldots < t_n$ and $i_0, i_1, \ldots, i_n \in E$ we have

$$\mathsf{P}[X_0 = i_0, X_{t_1} = i_1, \ldots, X_{t_n} = i_n] = p_{i_0}(0)p_{i_0 i_1}(t_1)p_{i_1 i_2}(t_2 - t_1)\ldots p_{i_{n-1} i_n}(t_n - t_{n-1}).$$

Also for a continuous–time Markov chain it is easy to verify the Chapman–Kolomogorov equations

$$\mathbf{P}(s)\mathbf{P}(t) = \mathbf{P}(s + t).$$

The transition probability matrix $\mathbf{P}(t)$ is differentiable with respect to t (see Kijima (1997, Theorem 4.4)) and since $\mathbf{P}(0) = \mathbf{I}$, we can compute the matrix $\mathbf{Q}$ as the right–side derivative of $\mathbf{P}(t)$ at point 0, i.e.,

$$\mathbf{Q} = \lim_{t \downarrow 0} \frac{\mathbf{P}(t) - \mathbf{I}}{t}.$$

The matrix $\mathbf{Q} = (q_{ij})_{i,j=1}^{K}$ is called the *infinitesimal generator*, or *generator* for short, of a finite continuous–time Markov chain. The generator matrix $\mathbf{Q}$ satisfies for every $i \in E$

$$\begin{aligned} q_{ij} &\geq 0, \qquad i \neq j, \\ q_{ii} &\leq 0, \\ q_i = -q_{ii} &= \sum_{j \neq i} q_{ij}, \end{aligned}$$

where q_{ij} are called *intensities* and q_i is often called the *total intensity*. If we have the generator $\mathbf{Q}$ we can get the transition probability matrix $\mathbf{P}(t)$ as the unique solution of the backward Kolmogorov differential equation

$$\mathrm{d}\mathbf{P}(t) = \mathbf{Q}\mathbf{P}(t)\,\mathrm{d}t, \qquad t \geq 0,$$

or the forward Kolmogorov differential equation

$$\mathrm{d}\mathbf{P}(t) = \mathbf{P}(t)\mathbf{Q}\,\mathrm{d}t, \qquad t \geq 0,$$

with the initial condition $\mathbf{P}(0) = \mathbf{I}$. The unique solution is given by

$$\mathbf{P}(t) = \exp(\mathbf{t}\mathbf{Q}) = \sum_{n=0}^{\infty} \frac{t^n \mathbf{Q}^n}{n!}, \qquad t \geq 0.$$

Let us denote the random time, when the process X_t changes the state for the i-th time, by T_i. We get an increasing sequence of stopping times $T_1 < T_2 < \ldots$. Times between jumps (changes of state) are random variables having an exponential distribution (see Kijima (1997, Theorem 4.6)). The intensity of the exponential distribution depends on the state in which the process X_t is. If the process X_t is in the state i, the random *sojourn time* before the next jump will occur is an exponentially distributed random variable τ_i with intensity q_i,

$$\mathsf{P}[\tau_i \leq t] = 1 - \mathrm{e}^{-q_i t}, \qquad t \geq 0. \tag{2.10}$$

In the case of zero intensity q_i the sojourn time is ∞ and the state i is absorbing, i.e., once the process X_t enters state i it will stay there forever.[10]

[10] If the state i is absorbing, the i-th row of the generator matrix is just the zero vector.

When the process X_t is in state i and a jump occurred, the probability that the next state will be state j is (for proof see Kijima (1997, Theorem 4.7))

$$\mathsf{P}[X_t = j | X_{t-} = i, X_t \neq i] = \frac{q_{ij}}{q_i}. \tag{2.11}$$

The probabilities (2.11) are transition probabilities of the embedded discrete–time Markov chain.

The continuous–time Markov chain can be defined through the embedded Markov chain and specifying intensities q_i for the distribution of the sojourn times. This definition will be handy for extending Markov chains to semi–Markov chains in Section 2.2.3.

Maximum Likelihood Estimation

Similar to the case of a discrete–time Markov chain we are interested in the maximum likelihood estimator of the generator matrix $\mathbf{Q}$ from observed data. If we have a continuous observation of the process X_t up to the present time T and we know, that exactly m jumps occurred, we can denote the jump times of the process by $J_1, J_2, \ldots, J_m$. It means that the process X_t has started in some initial state $i_0 \in E$, at time J_1 it has changed the state i_0 to state i_1 and so on until at time J_m it has jumped to state i_m, where it resides. Assuming we are given initial distribution we use (2.10) and (2.11) and get the maximum likelihood equation in form[11]

$$\begin{aligned}
L(\mathbf{Q}) &= q_{i_0} \exp(-q_{i_0} J_1) \\
&\quad \times \prod_{k=1}^{m-1} \frac{q_{i_{k-1} i_k}}{q_{i_{k-1}}} q_{i_k} \exp\big(-q_{i_k}(J_{k+1} - J_k)\big) \\
&\quad \times \frac{q_{i_{m-1} i_m}}{q_{i_{m-1}}} \exp\big(-q_{i_m}(T - J_m)\big) \\
&= \exp(-q_{i_0} J_1) \prod_{k=1}^{m-1} q_{i_{k-1} i_k} \exp\big(-q_{i_k}(J_{k+1} - J_k)\big) q_{i_{m-1} i_m} \exp\big(-q_{i_m}(T - J_m)\big) \\
&= \prod_{i=1}^{K} \exp\big(-q_i R_i\big) \prod_{\substack{j=1 \\ j \neq i}}^{K} q_{ij}^{n_{ij}} \\
&= \prod_{i=1}^{K} \exp\left(-\sum_{k \neq i} q_{ik} R_i\right) \prod_{\substack{j=1 \\ j \neq i}}^{K} q_{ij}^{n_{ij}} \\
&= \prod_{i=1}^{K} \prod_{\substack{j=1 \\ j \neq i}}^{K} q_{ij}^{n_{ij}} \exp\big(-q_{ij} R_i\big),
\end{aligned}$$

where n_{ij} is the total number of transitions from state i to state j up to time T, $R_i = \int_0^T \mathbf{1}_{\{x(t)=i\}} \, \mathrm{d}t$ is the total time spend by the process in state i up to time T and the term $\exp\big(-q_{i_m}(T - J_m)\big)$ is the probability that the last sojourn time is bigger than $T - J_m$. The

[11] In many publications we can find slightly different maximum likelihood function, which is derived from n independent runs of the system (instead of just one run). In that case the maximum likelihood function is a product of n functions in the form presented here. Estimation in that case boil down to same result only for derivation we need an additional index, we present ML estimator from one run. However, extension to n independent runs is obvious.

log–likelihood function is given by

$$l(\mathbf{Q}) = \log L(\mathbf{Q}) = \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \big(n_{ij} \log q_{ij} - R_i q_{ij} \big).$$

Let us differentiate $l(\mathbf{Q})$ with respect to q_{ij} and set the derivative equal to 0. We get

$$\frac{\partial}{\partial q_{ij}} l(\mathbf{Q}) = \frac{1}{q_{ij}} n_{ij} - R_i = 0, \qquad i \neq j.$$

The maximum likelihood estimator of the generator matrix $\mathbf{Q}$ is then

$$\begin{aligned} \hat{q}_{ij} &= \frac{n_{ij}}{R_i}, \qquad i \neq j, \\ \hat{q}_{ii} &= -\sum_{\substack{j=1 \\ j \neq i}}^{K} \hat{q}_{ij}. \end{aligned} \tag{2.12}$$

If we observe data continuously, the estimator (2.12) is useful. But if we observe the state of the process just in discrete equidistant times there is a problem to estimate the generator $\mathbf{Q}$. The way to deal with this, is described in Section 2.3.

Inhomogeneous Continuous–time Markov Chains

In the case of an inhomogeneous continuous–time Markov chain the generator $\mathbf{Q}$ is not constant anymore and it depends on time. For $0 \leq s < t$ we have

$$\begin{aligned} \mathbf{Q}(t) &= \lim_{h \downarrow 0} \frac{\mathbf{P}(t, t+h) - \mathbf{I}}{h}, \\ \mathbf{P}(s,t) &= \exp\left(\int_s^t \mathbf{Q}(u)\, \mathrm{d}u \right) = \sum_{n=0}^{\infty} \frac{\left(\int_s^t \mathbf{Q}(u)\, \mathrm{d}u \right)^n}{n!}, \end{aligned}$$

where the integral from a matrix is assumed to be componentwise, i.e.,

$$\int_s^t \mathbf{Q}(u)\, \mathrm{d}u = \begin{pmatrix} \int_s^t q_{1,1}(u)\, \mathrm{d}u & \int_s^t q_{1,2}(u)\, \mathrm{d}u & \cdots & \int_s^t q_{1,K}(u)\, \mathrm{d}u \\ \int_s^t q_{2,1}(u)\, \mathrm{d}u & \int_s^t q_{2,2}(u)\, \mathrm{d}u & \cdots & \int_s^t q_{2,K}(u)\, \mathrm{d}u \\ \vdots & \vdots & \ddots & \vdots \\ \int_s^t q_{K,1}(u)\, \mathrm{d}u & \int_s^t q_{K,2}(u)\, \mathrm{d}u & \cdots & \int_s^t q_{K,K}(u)\, \mathrm{d}u \end{pmatrix}.$$

For every $t \geq 0$ and $i \in E$ it holds

$$\begin{aligned} q_{ij}(t) &\geq 0, \qquad j \neq i, \\ q_{ii}(t) &\leq 0, \\ q_i(t) &= -q_{ii}(t) = \sum_{j \neq i} q_{ij}(t). \end{aligned}$$

2.2.3 Semi–Markov Chains

The continuous–time Markov chain is an extension of the discrete–time Markov chain when sojourn times are not a natural multiples of the time unit anymore, but they are replaced with an exponentially distributed random variable with a constant intensity. If we go further and replace the exponential distribution by any general distribution we will get a *semi–Markov chain* and if we allow a dependency of the distribution on time t we get an inhomogeneous semi–Markov chain. In this section we describe basic properties of finite inhomogeneous semi–Markov chains. Properties for homogeneous semi–Markov chains can be derived by relaxing the dependency on time t. Semi–Markov chains are often used in renewal theory and reliability theory. For the general theory of semi–Markov chains we refer to Janssen and Manca (2006).

Similar as in the case of a finite continuous–time Markov chain we have an embedded discrete–time Markov chain $\{J_n\}_{n\in\mathbb{N}_0}$ with state space $\{1,\dots,K\}$ and an increasing sequence of stopping times $\{T_n\}_{n\in\mathbb{N}}$. A semi–Markov chain X_t is fully described by (J_n, T_n) in the following way:

$$\begin{aligned} X_t &= J_0, \qquad t \in [0, T_1), \\ X_t &= J_1, \qquad t \in [T_1, T_2), \\ &\vdots \\ X_t &= J_n, \qquad t \in [T_n, T_{n+1}), \\ &\vdots \end{aligned}$$

Let us define a kernel associated to the semi–Markov chain X_t as

$$\mathbf{H}(s,t) = \Big(h_{ij}(s,t)\Big)_{i,j=1}^{K},$$

where

$$h_{ij}(s,t) = \mathsf{P}[J_{n+1} = j, T_{n+1} \le t | J_n = i, T_n = s].$$

Next, let us define

$$S_i(s,t) = \sum_{j=1}^{K} h_{ij}(s,t) = \mathsf{P}[T_{n+1} \le t | J_n = i, T_n = s],$$

that is, the probability distribution of the sojourn time in state i conditioned that the last jump occurred at time s. Finally, we can compute the transition probabilities $\mathbf{P}(s,t) = (p_{ij}(s,t))_{i,j}^{K}$ of the process X_t from evolution equations

$$p_{ij}(s,t) = \delta_{ij}\big(1 - S_i(s,t)\big) + \sum_{k=1}^{K} \int_s^t p_{kj}(u,t) h_{ik}(s,\, \mathrm{d}u), \tag{2.13}$$

where $\delta_{ij} = \mathbf{1}_{\{i=j\}}$ is the Kronecker delta. The general integral equations of form (2.13) are called Volterra integral equations. The equations (2.13) are extended Kolmogorov differential equations and can not be solved without any further specification of the sojourn time distribution and the inhomogeneity structure. If we assume some distribution, we can use the maximum likelihood estimation. If we do not want to assume any particular distribution we can use a non–parametric estimation, see Lucas et al. (2006). Working with a non–parametric estimation is harder than with a parametric estimation.

2.3 Partially Observed Data in Markov Chain Theory

Let us imagine the situation when we want to model some system using a continuous–time Markov chain. We have some past observations of the system from which we need to estimate the generator matrix $\mathbf{Q}$. If we have continuous–time observations of the system, it is easy to compute the maximum likelihood estimator. A problem rises up if we do not have continuous observations and we have just discretely (partially) observed data. Without loss of generality let us assume that we have yearly data observations.

Using the cohort method (2.9) we can estimate the yearly transition probability matrix and denote it by $\hat{\mathbf{P}}$. Then we would like to derive the estimator of the generator $\mathbf{Q}$ which generates the transition matrix $\hat{\mathbf{P}}$. This problem is called the *embeddability problem*. Not every probability matrix is embeddable and if there exists a generator $\mathbf{Q}$ it need not to be unique.

Under the assumption that the process can change the state just once a year Jarrow et al. (1997) show how to derive $\hat{\mathbf{Q}}$ from $\hat{\mathbf{P}}$. In Israel et al. (2001), the authors work in a more general setting without any restrictive assumptions and they provide sufficient conditions on the transition matrix $\hat{\mathbf{P}}$ for the existence or for the failure of the existence of the generator $\hat{\mathbf{Q}}$. Their idea is to compute the logarithm of matrix $\hat{\mathbf{P}}$. The resulting matrix does not need to be a valid generator.[12] Israel et al. (2001) come up with two ad–hoc methods how to get the nearest possible valid generator. Both methods set the negative off–diagonal elements equal to zero and then adjust other elements to get the zero row–sum. We describe them in Section 2.3.1

Another method how to get a valid generator matrix, which generates the transition probability matrix as near as possible to $\hat{\mathbf{P}}$, is from Kreinin and Sidelnikova (2001) (QO method) and described in Section 2.3.2.

In Bladt and Sørensen (2005) they choose a slightly different approach. They did not try to find the generator of the matrix $\hat{\mathbf{P}}$, but rather estimate the generator matrix $\mathbf{Q}$ using all available information, i.e., partially observed data. As a tool how to deal with the incomplete information for the maximum likelihood estimator, they used the EM algorithm (Section 2.3.3) and MCMC (Section 2.3.4).

A comprehensive comparison of these five methods can be found in Inamura (2006) who shows that the best performance is obtained by the MCMC method.

2.3.1 DA and WA Method

Israel et al. (2001) suggest to use the matrix logarithm through the Taylor series for computing the estimator

$$\hat{\mathbf{Q}} = \log \hat{\mathbf{P}} = (\hat{\mathbf{P}} - \mathbf{I}) - \frac{(\hat{\mathbf{P}} - \mathbf{I})^2}{2} + \frac{(\hat{\mathbf{P}} - \mathbf{I})^3}{3} - \frac{(\hat{\mathbf{P}} - \mathbf{I})^4}{4} + \dots,$$

where $\mathbf{I}$ is the identity matrix. Israel et al. (2001) also provide some necessary conditions for the existence of the real matrix logarithm of the matrix $\hat{\mathbf{P}}$. Nevertheless, even if the real logarithm exists, the resulting $\hat{\mathbf{Q}}$ does not need to be a valid generator. Israel et al. (2001) propose two ad–hoc methods how to fix it — the *diagonal adjustment method* (DA) and the *weighted adjustment method* (WA). The methods proceed in two steps. In the first step they fix all off–diagonal elements and in the second step they adjust other elements of $\hat{\mathbf{Q}}$ to be a valid generator.

[12] Non–positive elements on the diagonal, non–negative elements otherwise and zero row–sum.

1. Set all off–diagonal negative elements equal to 0,

$$\hat{q}_{ij} = \max\{\hat{q}_{ij}, 0\}, \qquad i \neq j.$$

2. For the DA method set the diagonal element equal to minus sum of off–diagonal elements

$$\hat{q}_{ii} = -\sum_{j \neq i} \hat{q}_{ij}, \qquad i = 1, \ldots, K.$$

For the WA method adjust all non–zero elements relative to their magnitudes

$$\hat{q}_{ij} = \hat{q}_{ij} - |\hat{q}_{ij}| \frac{\sum_{j=1}^{K} \hat{q}_{ij}}{\sum_{j=1}^{K} |\hat{q}_{ij}|}, \qquad i, j = 1, \ldots, K.$$

2.3.2 Quasi–optimization Method

We are given the matrix $\hat{\mathbf{P}}$ and we would like to find a generator which generates the matrix $\hat{\mathbf{P}}$ or generates a matrix which is as close as possible to $\hat{\mathbf{P}}$. Let us denote the set of all possible $K \times K$-dimensional generators by $\mathcal{Q}$, that is

$$\mathcal{Q} := \left\{ \mathbf{Q} \in \mathbb{R}^{K \times K} \middle| \sum_{j=1}^{K} q_{ij} = 0, q_{ij} \geq 0, i \neq j \right\}.$$

We are trying to find the element $\mathbf{Q}^* \in \mathcal{Q}$ that minimizes

$$||\exp(\mathbf{Q}) - \hat{\mathbf{P}}||,$$

where $\|\cdot\|$ is Euclidean norm. That is not an easy tractable problem. Kreinin and Sidelnikova (2001) suggest that if $\exp(\mathbf{Q})$ is near to $\hat{\mathbf{P}}$, then also their logarithms should be near to each other. Therefore, they formulate a similar problem, where they are looking for $\mathbf{Q}^* \in \mathcal{Q}$ that minimizes

$$||\mathbf{Q} - \log \hat{\mathbf{P}}||.$$

Since the conditions on a valid generator are conditions on every row we can split the optimization problem into K separate problems and perform for each row the following optimization algorithm.

Let us denote the particular row of $\hat{\mathbf{P}}$, for which we are performing the algorithm, by $\boldsymbol{a} = (a_1, \ldots, a_K)$ and assume that the cone $\mathcal{C}$ is given by

$$\mathcal{C} = \left\{ z \in \mathbb{R}^K \middle| \sum_{i=1}^{K} z_i = 0, \sum_{i=1}^{K} \mathbf{1}_{\{z_i < 0\}} \in \{0, 1\} \right\}.$$

Note that $\sum_{i=1}^{K} \mathbf{1}_{\{z_i<0\}} = 0$ implies $\boldsymbol{z} = \boldsymbol{0}$.

Now we are looking for $\boldsymbol{z}^* \in \mathcal{C}$ which is the argument of

$$\min_{z \in \mathcal{C}} \sum_{i=1}^{K} (a_i - z_i)^2.$$

Kreinin and Sidelnikova (2001) provide a fast algorithm for finding $\boldsymbol{z}^*$ based on the algorithm of Tuenter (2001). The algorithm proceeds as follows

1. If $\boldsymbol{a} \in \mathcal{C}$, then $\boldsymbol{z}^* = \boldsymbol{a}$ and stop.
2. Construct $\boldsymbol{b} = (b_1, \ldots, b_K)$ as
$$b_i = a_i - \frac{1}{K} \sum_{j=1}^{K} a_j.$$
3. Let $\pi(\cdot)$ be a permutation which orders $\boldsymbol{b}$ in the increasing order, that is $b_i \leq b_{i+1}$. Then compute $\tilde{\boldsymbol{a}} = \pi(\boldsymbol{b})$.
4. Find m^* as the smallest m, $1 \leq m \leq (K-1)$, which satisfies[13]
$$(K - m + 1)\tilde{a}_{m+1} - \left(\tilde{a}_1 + \sum_{i=m+1}^{K} \tilde{a}_i \right) \geq 0.$$
5. Let $\mathcal{J}$ be given by
$$\mathcal{J} := \{ i \in \mathbb{N} | 2 \leq i \leq m^* \}.$$
Note that $\mathcal{J}$ can be an empty set if $m^* = 1$.
6. Construct $\boldsymbol{z}^*$ by
$$z_i = \begin{cases} 0, & i \in \mathcal{J}, \\ \tilde{a}_i - \frac{1}{K - m^* + 1} \left(\sum_{j \notin \mathcal{J}} \tilde{a}_j \right), & \text{otherwise.} \end{cases}$$
Note that in the case of $m^* = 1$, we get simply $\boldsymbol{z}^* = \tilde{\boldsymbol{a}}$.
7. As the optimal solution for this row return $\pi^{-1}(\boldsymbol{z}^*)$, where π^{-1} is the inverse permutation to π.

If we perform the algorithm for every row, we get the solution to the QO problem.

2.3.3 Expectation Maximization Method

The idea of the expectation maximization method (EM) is quite simple and appears to be useful in the situation of incomplete data, in which the maximum likelihood could not be applied directly. For more general information about EM algorithm see McLachlan and Krishnan (1997). The application for the presented problem was proposed by Bladt and Sørensen (2005).

We assume that we have a discrete (one year) observation of n realizations of the continuous–time Markov chain. We can not apply the maximum likelihood since we do not observe these realizations continuously, but we know in which state the chain started and where it ended. For computing the maximum likelihood we would need to know the number of transitions from i to j through the year denoted by n_{ij} and the time spend in state i by all realizations denoted by R_i.

The EM algorithm consists of two steps. In the first step, n_{ij} and R_i are replaced by their expected values, given the partial observation $\boldsymbol{x}^{obs}$ (starting and ending point). In the second step, we compute the maximum likelihood estimator of the generator $\mathbf{Q}$ using n_{ij} and R_i from the first step. Formally the algorithm proceeds as follows

[13] In the original paper Kreinin and Sidelnikova (2001) or in Inamura (2006) they request $m \geq 2$, which we think is wrong and does not work in some special cases (matrices in credit risk usually do not satisfy this). Paper of Kreinin and Sidelnikova (2001) is based on paper Tuenter (2001). Tuenter (2001) proved his algorithm for a slightly different simplex, so the mistake probably occurred during adjustments for different simplex.

1. Let $\mathbf{Q}_0$ be some starting point of the algorithm. The possible choice is the estimator $\mathbf{Q}_{DA}$ or $\mathbf{Q}_{WA}$.
2. Calculate $\mathsf{E}\,[n_{ij}|\boldsymbol{x}^{obs}, \mathbf{Q}_k]$ and $\mathsf{E}\,[R_i|\boldsymbol{x}^{obs}, \mathbf{Q}_k]$ using $\mathbf{Q}_k$.
3. Calculate $\mathbf{Q}_{k+1}$, where the ij-th element is given by

$$q_{ij} = \frac{\mathsf{E}\,[n_{ij}|\boldsymbol{x}^{obs}, \mathbf{Q}_k]}{\mathsf{E}\,[R_i|\boldsymbol{x}^{obs}, \mathbf{Q}_k]}.$$

4. If $||\mathbf{Q}_k - \mathbf{Q}_{k+1}|| < \varepsilon$, return $\mathbf{Q}_k$ as the optimal solution, else go to step 2.

The only thing we need to clarify is how to compute expectations $\mathsf{E}\,[n_{ij}|\boldsymbol{x}^{obs}, \mathbf{Q}]$ and $\mathsf{E}\,[R_i|\boldsymbol{x}^{obs}, \mathbf{Q}]$. Note that

$$\mathsf{E}\,[n_{ij}|\boldsymbol{x}^{obs}, \mathbf{Q}] = \sum_{h=1}^{n} \mathsf{E}\,\left[n_{ij}^h|x^h, \mathbf{Q}\right], \tag{2.14}$$

$$\mathsf{E}\,[R_i|\boldsymbol{x}^{obs}, \mathbf{Q}] = \sum_{h=1}^{n} \mathsf{E}\,\left[R_i^h|x^h, \mathbf{Q}\right], \tag{2.15}$$

where x^h is the partial observation of the h-th realization of the chain, n_{ij}^h is the number of transitions from state i to state j in the h-th realization and similar to R_i^h. Then it holds

$$\mathsf{E}\,\left[n_{ij}^h|x^h, \mathbf{Q}\right] = \frac{1}{D}\boldsymbol{e}_{x^h(0)}^\top \left(q_{ij}\int_0^1 \exp(s\mathbf{Q})\boldsymbol{e}_i\boldsymbol{e}_j^\top \exp\big((1-s)\mathbf{Q}\big)\,\mathrm{d}s\right)\boldsymbol{e}_{x^h(1)}, \tag{2.16}$$

$$\mathsf{E}\,\left[R_i^h|x^h, \mathbf{Q}\right] = \frac{1}{D}\boldsymbol{e}_{x^h(0)}^\top \left(\int_0^1 \exp(s\mathbf{Q})\boldsymbol{e}_i\boldsymbol{e}_i^\top \exp\big((1-s)\mathbf{Q}\big)\,\mathrm{d}s\right)\boldsymbol{e}_{x^h(1)}, \tag{2.17}$$

$$D = \boldsymbol{e}_{x^h(0)}^\top \exp(\mathbf{Q})\boldsymbol{e}_{x^h(1)}, \tag{2.18}$$

where $x^h(0)$ is the starting and $x^h(1)$ the ending state of the h-th realization, $\boldsymbol{e}_i$ is a vector of zeros with one on i-th place. For the detailed derivation of formulas (2.16) – (2.18), see Bladt and Sørensen (2005) and/or Inamura (2006, Appendix B). To be able to compute any of the formulas (2.16) – (2.18), we need to know how to compute integrals involving matrix exponentials. An easy formula suggested by Van Loan (1978) use two square matrices $\mathbf{F}$ and $\mathbf{A}$ such that

$$\begin{pmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{0} & \mathbf{F}_{22} \end{pmatrix} = \exp\left[t\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{pmatrix}\right].$$

Then it holds that

$$\begin{aligned}
\mathbf{F}_{11} &= \exp(t\mathbf{A}_{11}),\\
\mathbf{F}_{22} &= \exp(t\mathbf{A}_{22}),\\
\mathbf{F}_{12} &= \int_0^t \exp\big((t-s)\mathbf{A}_{11}\big)\mathbf{A}_{12}\exp(s\mathbf{A}_{22})\,\mathrm{d}s.
\end{aligned}$$

If we replace $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ by $\mathbf{Q}$, $\mathbf{A}_{12}$ by $\boldsymbol{e}_i\boldsymbol{e}_i^\top$ or $\boldsymbol{e}_i\boldsymbol{e}_j^\top$ and $t = 1$, we can easily compute $\exp(\mathbf{A})$ and get $\mathbf{F}_{12}$, which is what we are looking for.

2.3.4 Markov Chain Monte Carlo Method

The last method from the comparison of Inamura (2006) is the MCMC method proposed by Bladt and Sørensen (2005). The idea is also very simple. Let $\mathbf{Q}$ be a random matrix satisfying conditions on the generator matrix with the distribution $p(\mathbf{Q})$. Let us draw some particular $\mathbf{Q}$ from $p(\mathbf{Q})$ and simulate the run X of the chain given the generator matrix $\mathbf{Q}$ in a such way, that the partial observation (starting and ending state) agrees with data which we have observed $\boldsymbol{x}^{obs}$. Then we compute the maximum likelihood from the run X and adjust the distribution of $\mathbf{Q}$ according to that. The setting is Bayesian and from the Bayes theorem we have

$$\begin{aligned} p(\mathbf{Q}|X, \boldsymbol{x}^{obs}) &\propto p(\mathbf{Q})p(X|\mathbf{Q}) \\ &\propto p(\mathbf{Q})L(\mathbf{Q}) \\ &= p(\mathbf{Q}) \prod_{i=1}^{K} \prod_{\substack{j=1 \\ j\neq i}}^{K} q_{ij}^{n_{ij}} \exp(-q_{ij}R_i). \end{aligned} \tag{2.19}$$

Thus, $p(\mathbf{Q})$ is the prior distribution and $p(\mathbf{Q}|X, \boldsymbol{x}^{obs})$ is the posterior distribution. This distribution adjustment of the matrix $\mathbf{Q}$, we iterate n-times and after that we forget the first l iterations, which is called the *burn–in period*. Then we get the estimator

$$\hat{\mathbf{Q}} = \frac{1}{n-l} \sum_{i=l+1}^{n} \mathbf{Q}_i,$$

where $\mathbf{Q}_i$ is the matrix drawn in the i-th iteration. The first l iterations are discarded to enable the process to reach stationarity. Bladt and Sørensen (2005) proposed to use a gamma distribution for off–diagonal elements of $\mathbf{Q}$ (diagonal elements are given by $q_{ii} = -\sum_{j\neq i} q_{ij}$). Then the prior distribution is

$$p(\mathbf{Q}) \propto \prod_{i=1}^{K} \prod_{\substack{j=1 \\ j\neq i}}^{K} q_{ij}^{\alpha_{ij}-1} \exp(-q_{ij}\beta_i),$$

where α_{ij} and β_i are parameters of the gamma distribution. The posterior distribution is then from (2.19) again the gamma distribution, but with different parameters

$$p(\mathbf{Q}|X, \boldsymbol{x}^{obs}) \propto \prod_{i=1}^{K} \prod_{\substack{j=1 \\ j\neq i}}^{K} q_{ij}^{n_{ij}+\alpha_{ij}-1} \exp\big(-q_{ij}(R_i + \beta_i)\big).$$

Let α_{ij} and β_i be given parameters of the gamma distribution and $\boldsymbol{x}^{obs}$ the partial observation. The algorithm proceeds as follows:

1. Let $k = 1$.
2. Draw $\mathbf{Q}_k$ from distribution $\Gamma(\boldsymbol{\alpha}, \boldsymbol{\beta})$.
3. Simulate the run X of the chain in such way that X agrees with the partial observation $\boldsymbol{x}^{obs}$.
4. Compute n_{ij} and R_i from run X and update α_{ij} and β_i,

$$\begin{aligned} \alpha_{ij} &= \alpha_{ij} + n_{ij}, \\ \beta_i &= \beta_i + R_i, \\ k &= k + 1. \end{aligned}$$

5. If $k = (n+1)$ return $\frac{1}{n-l}\sum_{i=l+1}^{n}\mathbf{Q}_i$, else go to step 2.

The simulation of run X can be done easily by simulating a Markov chain with the initial state according to $\boldsymbol{x}^{obs}$. Then the run is accepted or rejected if it will end in the desired ending point (according to $\boldsymbol{x}^{obs}$).

Note that the generator estimators from the EM and the MCMC method can generate significantly different one–year transition matrices than $\hat{\mathbf{P}}$. It is caused by a similar reason as in the case of the ML estimator compared with the cohort estimator. The EM and the MCMC method incorporate more information than the methods DA, WA and QO. Therefore, it is not easy to compare these two kinds of methods. Sometimes the first one might be more useful than the other one dependent on the observed information.

Inamura (2006) performed a simulation when he has some generator $\mathbf{Q}$ and he simulates the partial observation and compares estimators from these five methods with the original one. The best estimator was given by the MCMC method. The QO method gives a better fit if one is looking for the generator of $\hat{\mathbf{P}}$.

2.3.5 Componentwise Optimization Method

As an additional method to the five previous ones we add the componentwise optimization (CO), which has a similar goal as the DA, WA, or QO method — find $\mathbf{Q}$ which generates $\hat{\mathbf{P}}$. The idea is very simple — divide the problem of finding $\mathbf{Q}$ into $(K-1)\times(K-1)$ steps. In each step we are dealing with a one–dimensional problem, where we find

$$\hat{q}_{ij} = \arg\min_{q_{ij}\in[0,c]} \left\| \exp\left(\mathbf{Q}(q_{ij})\right) - \hat{\mathbf{P}} \right\|, \qquad i \neq j, \tag{2.20}$$

where $\mathbf{Q}(q_{ij})$ expresses the dependency of the generator on q_{ij}. The constant c is chosen in a conservative way from the empirical expectations to get a bounded problem.[14] The smaller the constant c is, the faster the algorithm proceeds. The constant c can be also different for different choices of indices i and j.

We see that in each step we have fixed all elements of $\mathbf{Q}$ except particular $q_{ij}, i \neq j$, and we perform a one–dimensional optimization. The important thing is, that simultaneously as we are moving q_{ij}, we need to adjust other elements of the i-th row of $\mathbf{Q}$ to remain a valid generator. This one dimensional problem is solvable numerically using a wide variety of methods.[15]

Formally the algorithm is as follows:

1. Find some initial $\mathbf{Q}_0$ using the DA or the WA method.
2. While convergence is not reached do
   - For $i = 1, \ldots, K$ do
     - For $j = 1, \ldots, K$ do
       - If $i \neq j$, find $\hat{q}_{ij}$ as

$$\hat{q}_{ij} = \arg\min_{q_{ij}\in[0,c]} \left\| \exp\left(\mathbf{Q}(q_{ij})\right) - \hat{\mathbf{P}} \right\|,$$

         Simultaneously with changing q_{ij} adjust other row elements to $\mathbf{Q}$ remain a valid generator.

[14] In the credit risk, the highest intensities are around 0.1, therefore a possible conservative choice would be $c = 1$.

[15] We have used the function `fminbnd` from the mathematical software Matlab, which find a minimum on bounded interval. The algorithm is based on golden section search and parabolic interpolation. For details see Brent (1973).

3. Return **Q** as the optimal solution.

The CO algorithm does not need to lead to the optimal solution, since it converges to some local minimum of one dimension and since we do not know anything about convexity, we can not ensure that the found minimum is a global minimum. Hence, we know, that we will get "some" result, but the result can be far away from the optimal solution. What we can do, is to compare it to solutions from the previous five methods and then say if it is better solution or it is not. The important thing is also the choice of the initial point since we never converge to a worse solution than the initial. Hence, if we choose as a starting point the best solution which we obtained from the previous five methods, we can just improve using CO. The standalone usage without any reference results can not be done without further discussion about the convexity of the problem.

3
Time–changed Continuous–time Markov Chains

This chapter is a main contribution of the presented work. The motivation for our extension is following. Let us consider some system which is described by a finite homogeneous continuous–time Markov chain X_t. The homogeneity assumption is often very controversial. Empirical studies often show the violence of system time homogeneity. When we want to fix this flaw, we need to introduce a more general system (inhomogeneous (semi–) Markov chain). In the inhomogeneity structure we would like to conserve as fewest parameters as possible, since every new parameter in model will raise data demand needed for estimation. One of possible suggestions for the inhomogeneity structure can be found in Bluhm and Overbeck (2007).

We propose a different way inspired mainly by time changes in Lévy processes (see Carr and Wu (2004)). We assume that system follows a continuous–time Markov chain (CTMC), which is time–homogeneous with respect to a different time scale than a usual calendar time. In every field of application, this alternative time scale can have different interpretation.

Let us call the new time scale *system time* (or in a context with credit risk *business time*) and denote the mapping which maps regular calendar time into the system time by T_t. This mapping can be deterministic or random. We can treat the deterministic mapping as a special (degenerate) case of random mapping. Therefore, we are going to describe the theory for random T_t. Since T_t is a random variable for every t, we require the following properties

$$\begin{aligned} \mathsf{P}[T_0 = 0] &= 1, \\ \mathsf{P}[T_t < \infty] &= 1, \qquad t \geq 0, \\ \mathsf{P}[T_t \geq T_s] &= 1, \qquad t \geq s \geq 0, \\ \mathsf{P}[T_t > T_s] &> 0, \qquad t > s \geq 0. \end{aligned}$$

In other words, the process T_t is almost surely non–decreasing, starting at zero and its increments are with positive probability bigger than zero. Then we study the behavior of *time–changed continuous–time Markov chain*. The time evolution can be deterministic — given exogenously by another model or by expert opinion. Random time evolution is usually modeled using Lévy subordinators (see Section 2.1.4).

In Section 3.1 we extend classical theory and describe behavior of a time–changed CTMC. In Sections 3.2 and 3.3 we study how to estimate parameters in our model under different settings and approaches to time evolution.

3.1 Time–changed Continuous–time Markov Chains

Let us recall a few known results from matrix algebra, which we will need later. Then we will show how a continuous–time Markov chain evolves after the time change.

3.1.1 Matrix Exponential

Let us begin with the known fact that for any given $n \times n$ matrix $\mathbf{Q}$, over an algebraically closed field,[16] there exists an invertible $n \times n$ matrix $\mathbf{B}$ such that

$$\mathbf{Q} = \mathbf{B}\mathbf{D}\mathbf{B}^{-1}, \tag{3.1}$$

where $\mathbf{D}$ is a $n \times n$ block diagonal matrix given by

$$\mathbf{D} = \begin{pmatrix} \mathbf{J}_{n_1}(\lambda_1) & & & \mathbf{0} \\ & \mathbf{J}_{n_2}(\lambda_2) & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{J}_{n_k}(\lambda_k) \end{pmatrix},$$

and the block $\mathbf{J}_{n_i}(\lambda_i)$ is a $n_i \times n_i$ square matrix given in form

$$\mathbf{J}_{n_i}(\lambda_i) = \begin{pmatrix} \lambda_i & 1 & \dots & 0 \\ 0 & \lambda_i & \ddots & \vdots \\ \vdots & & \ddots & 1 \\ 0 & \dots & 0 & \lambda_i \end{pmatrix},$$

such that $n_1 + n_2 + \ldots + n_k = n$. The values $\lambda_1, \ldots, \lambda_k$ are eigenvalues of the matrix $\mathbf{Q}$. Form (3.1) is called the *Jordan canonical representation* and $\mathbf{J}_{n_i}(\lambda_i)$ is a *Jordan block* belonging to the eigenvalue λ_i. In the case $n = k$, that is, the block $\mathbf{J}_{n_i}(\lambda_i)$ is 1-dimensional for every i, we say, that the matrix $\mathbf{Q}$ is *diagonalizable*.

Let $\mathbf{J}_n(\lambda)$ be an arbitrary Jordan block. It can be decomposed into

$$\mathbf{J}_n(\lambda) = \lambda\mathbf{I} + \mathbf{N},$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{N}$ is a nilpotent matrix[17] in form

$$\mathbf{N} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & \ddots & \\ \vdots & \vdots & & \ddots & 1 \\ 0 & 0 & & & 0 \end{pmatrix}.$$

By easy multiplication we see that

$$\mathbf{N}^2 = \begin{pmatrix} 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ & & & \ddots & \vdots \\ & & & & 1 \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & & & 0 \end{pmatrix}, \dots, \mathbf{N}^{n-1} = \begin{pmatrix} 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \\ & & & \ddots & \vdots \\ & & & & 0 \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & & & 0 \end{pmatrix},$$

i.e., with higher powers of the matrix $\mathbf{N}$ the line of ones shifts one entry to the right. Hence the order of nilpotency of the matrix $\mathbf{N}$ is $n - 1$.

[16] $\mathbb{R}$ is not closed, but $\mathbb{C}$ is.

[17] We say that the matrix $\mathbf{N}$ is nilpotent of order $a \in \mathbb{N}$ if $\mathbf{N}^k \neq \mathbf{0}$ for $k < a$ and $\mathbf{N}^a = \mathbf{0}$.

It is easy to verify that matrices $\lambda\mathbf{I}$ and $\mathbf{N}$ commute. Then it holds that

$$\begin{aligned}
\mathrm{e}^{\lambda\mathbf{I}+\mathbf{N}} &= \sum_{i=0}^{\infty} \frac{(\lambda\mathbf{I}+\mathbf{N})^i}{i!} \\
&= \sum_{i=0}^{\infty} \frac{\sum_{j=0}^{i} \binom{i}{j} (\lambda\mathbf{I})^j \mathbf{N}^{i-j}}{i!} \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{\lambda^j \mathbf{N}^{i-j}}{j!(i-j)!} \\
&= \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \sum_{i=j}^{\infty} \frac{\mathbf{N}^{i-j}}{(i-j)!} \\
&= \mathrm{e}^{\lambda}\mathrm{e}^{\mathbf{N}}.
\end{aligned}$$

Since the matrix $\mathbf{N}$ is nilpotent of order $n-1$, the Taylor expansion of $\mathrm{e}^{\mathbf{N}}$ is a finite sum consisting of n terms. The first term of the expansion is the identity matrix and every other term shifts the ones on the diagonal one entry to the right as described above. We see that

$$\mathrm{e}^{\mathbf{J}_n(\lambda)} = \mathrm{e}^{\lambda\mathbf{I}+\mathbf{N}} = \mathrm{e}^{\lambda} \begin{pmatrix} 1 & 1 & \frac{1}{2} & \dots & \frac{1}{(n-1)!} \\ 0 & 1 & 1 & \dots & \frac{1}{(n-2)!} \\ 0 & 0 & 1 & \dots & \frac{1}{(n-3)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}. \tag{3.2}$$

Now we can compute the exponential of an arbitrary matrix $\mathbf{Q}$. Notice that

$$\mathrm{e}^{\mathbf{Q}} = \sum_{i=0}^{\infty} \frac{\left(\mathbf{B}\mathbf{D}\mathbf{B}^{-1}\right)^i}{i!} = \mathbf{B} \sum_{i=0}^{\infty} \frac{\mathbf{D}^i}{i!} \mathbf{B}^{-1} = \mathbf{B}\mathrm{e}^{\mathbf{D}}\mathbf{B}^{-1}.$$

We need to compute the exponential of the block diagonal matrix $\mathbf{D}$. Since the exponent of a block diagonal matrix is a block diagonal matrix with exponentials of original blocks on the diagonal, we have

$$\mathrm{e}^{\mathbf{D}} = \exp\left[\begin{pmatrix} \mathbf{J}_{n_1}(\lambda_1) & & & \mathbf{0} \\ & \mathbf{J}_{n_2}(\lambda_2) & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{J}_{n_k}(\lambda_k) \end{pmatrix}\right] = \begin{pmatrix} \mathrm{e}^{\mathbf{J}_{n_1}(\lambda_1)} & & & \mathbf{0} \\ & \mathrm{e}^{\mathbf{J}_{n_2}(\lambda_2)} & & \\ & & \ddots & \\ \mathbf{0} & & & \mathrm{e}^{\mathbf{J}_{n_k}(\lambda_k)} \end{pmatrix},$$

where exponentials of Jordan blocks are given by (3.2).

3.1.2 Continuous–time Markov Chains under Stochastic Time

Recall that transition probabilities of a homogeneous continuous–time Markov chain are given by

$$\mathbf{P}(t) = \exp\left(t\mathbf{Q}\right).$$

Let us try to move to stochastic time, where the time evolution is modeled through a non–decreasing stochastic process T_t. Conditioned on the realization T_t the transition probability matrix is

$$\mathbf{P}_{T_t}(t) = \exp\left(T_t\mathbf{Q}\right).$$

The unconditioned transition probability matrix is then

$$\mathbf{P}(t) = \mathsf{E}\left[\exp\left(T_t\mathbf{Q}\right)\right].$$

Let $\mathbf{Q} = \mathbf{B}\mathbf{D}\mathbf{B}^{-1}$ be the Jordan decomposition of the matrix $\mathbf{Q}$, then we have from the previous subsection

$$\mathbf{P}(t) = \mathsf{E}\left[\exp\left(T_t\mathbf{Q}\right)\right] = \mathbf{B}\mathsf{E}\left[\exp\left(T_t\mathbf{D}\right)\right]\mathbf{B}^{-1}.$$

We need to compute the expectation

$$\mathsf{E}\left[\exp\left(T_t\mathbf{D}\right)\right] = \mathsf{E}\left[\begin{pmatrix} \exp\left(T_t\mathbf{J}_{n_1}(\lambda_1)\right) & 0 & \cdots & 0 \\ 0 & \exp\left(T_t\mathbf{J}_{n_2}(\lambda_2)\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \exp\left(T_t\mathbf{J}_{n_k}(\lambda_k)\right) \end{pmatrix}\right].$$

Now, we need to investigate what has changed in the computation of the exponential of a Jordan block by moving to the stochastic time in comparison with (3.2),

$$\mathrm{e}^{T_t\mathbf{J}_n(\lambda)} = \mathrm{e}^{T_t\lambda\mathbf{I}+T_t\mathbf{N}} = \mathrm{e}^{T_t\lambda}\begin{pmatrix} 1 & T_t & \frac{T_t^2}{2} & \dots & \frac{T_t^{n-1}}{(n-1)!} \\ 0 & 1 & T_t & \dots & \frac{T_t^{n-2}}{(n-2)!} \\ 0 & 0 & 1 & \dots & \frac{T_t^{n-3}}{(n-3)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Note that the powers of T_t in the exponential matrix $\mathrm{e}^{T_t\mathbf{N}}$ came from the Taylor series. We see that we need to compute

$$\mathsf{E}\left[T_t^i\mathrm{e}^{T_t\lambda}\right], \qquad i = 0, \dots, n-1,$$

if the expectations exist. For $i = 0$ the question of expectation existence reduce to existence of Laplace transform.

If we model stochastic time evolution using Lévy subordinator expectation exists always for $i = 0$ since λ is non–positive[18]. For $i > 0$ we need to know the particular distribution of the Lévy process.

Case of Time–homogeneous Lévy Subordinator and Diagonizable Generator Matrix

In the special case when we assume the time–homogeneous Lévy subordinator T_t and the matrix $\mathbf{Q}$ is diagonalizable (the matrix $\mathbf{D}$ is diagonal), the Jordan blocks are one–dimensional and equal to

$$\mathsf{E}\left[\mathrm{e}^{T_t\mathbf{J}_n(\lambda)}\right] = \mathsf{E}\left[\mathrm{e}^{T_t\lambda}\right] = \mathrm{e}^{t\eta(\lambda)},$$

[18]The eigenvalue of the generator matrix $\mathbf{Q}$ with the biggest real part, is unique and equals 0. The rest of the eigenvalues have a negative real part. A proof can be found for example in Dupač and Dupačová (1980, Page 43).

where $\eta(u)$ is the Laplace exponent of the subordinator T_t. Let $\mathbf{E}$ be

$$\mathbf{E} = \begin{pmatrix} \eta(D_{11}) & 0 & \cdots & 0 \\ 0 & \eta(D_{22}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \eta(D_{KK}) \end{pmatrix},$$

then it holds that

$$\mathsf{E}\left[\exp\left(T_t\mathbf{D}\right)\right] = \exp\left(t\mathbf{E}\right),$$

and the transition probabilities are given by

$$\mathbf{P}(t) = \mathbf{B}\exp\left(t\mathbf{E}\right)\mathbf{B}^{-1} = \exp\left(t\mathbf{B}\mathbf{E}\mathbf{B}^{-1}\right) = \exp(t\tilde{\mathbf{Q}}),$$

where $\tilde{\mathbf{Q}} = \mathbf{B}\mathbf{E}\mathbf{B}^{-1}$. We see that we get a homogeneous continuous–time Markov chain with a different generator matrix. However, it is only the special case, if we have a time inhomogeneous subordinator or at least one Jordan block with dimension higher than 1 we can get a time–inhomogeneous continuous–time Markov chain (the generator $\mathbf{Q}_t$ will depend on time t.

3.2 Maximum Likelihood Estimation

Similar to the case of regular continuous–time Markov chain we would like to be able to estimate parameters of the model. In practice usually we are interested in system behavior in some given periods. We assume without loss of generality that this period is one year and hence we model yearly time increments.

Let us have observation of the system for m years. We assume that in year i system follows continuous–time Markov chain with generator matrix $t_i\mathbf{Q}$. We need to distinguish if the time evolution is deterministic or random. In the case of deterministic time evolution we need to further distinguish if time evolution is given or it has to be estimated as well. So we can look on the system in three possible settings which differ in the set of parameter which needs to be estimated.

TEEG – time evolution $t_1, \ldots, t_m$ is exogenously given and we need to estimate just the generator matrix $\mathbf{Q}$.

TEPG – time evolution $t_1, \ldots, t_m$ are parameters of the model and need to be estimated together with the generator matrix $\mathbf{Q}$.

TERV – time evolution $t_1, \ldots, t_m$ is assumed to be i.i.d. random variables from the probability distribution, which depends on parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$. In this case we need to estimate the generator matrix $\mathbf{Q}$ and parameters $\boldsymbol{\alpha}$.

There is one important thing which we need to take care of in the TEPG and TERV settings. Note that in the formula for the transition probability matrix, there is a term

$$\exp(t_k\mathbf{Q}),$$

where t_k is the time evolution realization in year k. In the TEPG settings t_k needs to be estimated as a parameter and in the TERV settings parameter of distribution of yearly increments in year k has to be estimated. However in both cases we can take any $c > 0$ and do the following

$$\mathrm{e}^{-t_k\mathbf{Q}} = \mathrm{e}^{-ct_k\frac{\mathbf{Q}}{c}} = \mathrm{e}^{-\tilde{t_k}\tilde{\mathbf{Q}}}.$$

The positive multiple of the time evolution t_k is also some time evolution and the positive multiple of the generator matrix $\mathbf{Q}$ is also the generator matrix. If we want to derive the unique maximum likelihood (ML) estimator we need to add some additional conditions.

In the TEPG settings seems reasonable condition

$$\sum_{i=1}^{m} t_i = m. \tag{3.3}$$

By which we force the same total time under both time scales (system time and calendar time).

In the TERV settings there is no such a general obvious choice. If the distribution of yearly increments depends on parameters $\alpha_1, \ldots, \alpha_n$, we can set α_1 equal to some constant c, but we need to ensure that all parameters in the model will have valid values. Particular choices differ from distribution to distribution.

In each of these cases (TEEG, TEPG and TERV) we are going to derive the maximum likelihood estimator. Since inter–arrival times of continuous–time Markov chain are exponentially distributed we need to know how the exponential distribution behaves under the time change, so in all cases we study the time–changed exponential distribution and later on, the time–changed continuous–time Markov chain.

3.2.1 Time Evolution Exogenously Given

Let us assume interval $(0, n]$ where n is an integer (number of years). Further we assume that we are given the time evolution during these n years $t_1, t_2, \ldots t_n$. The time evolution is described by a broken line. Further, let us assume interval $(c, T^*] \in (0, n]$. We are interested in a random time X until some event happen. This random time lives on interval (c, ∞). We are interested only in fact when it happened — if it happened before T^*. Under the calendar time X would be exponentially distributed and we would not care when it started (since memoryless of the exponential distribution).

We study the random variable $X : \Omega \to (c, \infty]$ with a following property. Let $k \in \{\lceil c \rceil, \ldots, n-1\}$. For every x such that $k < x \leq k+1$ holds

$$\mathsf{P}[X \leq x | X > k] = 1 - \exp\big(-\lambda t_{k+1}(x-k)\big),$$

where $\lambda > 0$ is some positive parameter. In other words we study a random variable which, conditioned on being bigger than k, behaves on interval $(k, k+1]$ as an exponentially distributed random variable with the intensity λt_{k+1}.

Let us derive the distribution function of X on interval $(c, n]$. We are not interested in the behavior of X in the region (n, ∞). For simplicity let us assume that $c < 1$ and $n > 2$.

For $c < x \leq 1$ we have from definition

$$\mathsf{P}[X \leq x] = 1 - \exp\big(-\lambda t_1(x-c)\big).$$

Let $1 < x \leq 2$, then

$$\begin{aligned}
\mathsf{P}[X \leq x] &= \mathsf{P}[X \leq 1] + \mathsf{P}[1 < X \leq x] \\
&= \mathsf{P}[X \leq 1] + \mathsf{P}[X \leq x | X > 1]\mathsf{P}[X > 1] \\
&= 1 - \exp\big(-\lambda t_1(1-c)\big) + \big(1 - \exp\big(-\lambda t_2(x-1)\big)\big)\exp\big(-\lambda t_1(1-c)\big) \\
&= 1 - \exp\Big(-\lambda\big(t_1(1-c) + t_2(x-1)\big)\Big).
\end{aligned}$$

By induction it easily follows that for $x \in (k, k+1]$, $n > k > 1$

$$\mathsf{P}[X \le x] = 1 - \exp\left(-\lambda\left(t_1(1-c) + \sum_{i=2}^{k} t_i + t_{k+1}(x-k)\right)\right).$$

If we differentiate the distribution function on interval $(k, k+1]$ we will get the density function

$$f_X(x) = \lambda t_{k+1} \exp\left(-\lambda\left(t_1(1-c) + \sum_{i=2}^{k} t_i + t_{k+1}(x-k)\right)\right).$$

For a general n and $c \in (0, n]$ we just need to play a bit with indices

$$\mathsf{P}[X \le x] = 1 - \exp\big(-\lambda D(c,x)\big), \qquad x \in (c, n],$$

where

$$\begin{aligned} D(c,x) &= t_{\lfloor x\rfloor+1}(x-c), && \lfloor x\rfloor - \lceil c\rceil = -1,\\ &= t_{\lceil c\rceil}(\lceil c\rceil - c) + t_{\lfloor x\rfloor+1}(x - \lfloor x\rfloor), && \lfloor x\rfloor - \lceil c\rceil = 0,\\ &= t_{\lceil c\rceil}(\lceil c\rceil - c) + \sum_{i=\lceil c\rceil+1}^{\lfloor x\rfloor} t_i + t_{\lfloor x\rfloor+1}(x - \lfloor x\rfloor), && \lfloor x\rfloor - \lceil c\rceil > 0, \end{aligned}$$

where $\lfloor\cdot\rfloor$ ($\lceil\cdot\rceil$) means lower (upper) integer part. The density function is then in form

$$f_X(x) = \lambda t_{\lfloor x\rfloor+1} \exp\left(-\lambda D(c,x)\right).$$

Let us study how the maximum likelihood estimator for CTMC parameters will look like. Let $J_0 = 0$ and $J_{M+1} = T^*$ and recall (see Section 2.2.2) that the usual maximum likelihood function is (assuming we are given the initial state i_0)

$$\begin{aligned} L(\mathbf{Q}) &= q_{i_0} \exp(-q_{i_0} J_1)\\ &\quad \times \prod_{k=1}^{m-1} \frac{q_{i_{k-1} i_k}}{q_{i_{k-1}}} q_{i_k} \exp\big(-q_{i_k}(J_{k+1} - J_k)\big)\\ &\quad \times \frac{q_{i_{m-1} i_m}}{q_{i_{m-1}}} \exp\big(-q_{i_m}(T^* - J_m)\big)\\ &= \prod_{i=1}^{K} \prod_{\substack{j=1\\ j\ne i}}^{K} q_{ij}^{n_{ij}} \exp\big(-q_{ij} R_i\big), \end{aligned}$$

where

$$R_i = \int_0^T \mathbf{1}_{\{x_t = i\}}\, \mathrm{d}t = \sum_{k=0}^{m} (J_{k+1} - J_k) \mathbf{1}_{\{i_k = i\}}.$$

Under the given time evolution we get the likelihood function

$$\begin{aligned}
L(\mathbf{Q}) &= q_{i_0} t_{\lfloor J_1 \rfloor + 1} \exp\big(- q_{i_0} D(0, J_1)\big) \\
&\quad \times \prod_{k=1}^{m-1} \frac{q_{i_{k-1} i_k}}{q_{i_{k-1}}} q_{i_k} t_{\lfloor J_{k+1} \rfloor + 1} \exp\big(- q_{i_k} D(J_k, J_{k+1})\big) \\
&\quad \times \frac{q_{i_{m-1} i_m}}{q_{i_{m-1}}} \exp\big(- q_{i_k} D(J_m, T^*)\big) \\
&= t_{\lfloor J_1 \rfloor + 1} \exp\big(- q_{i_0} D(0, J_1)\big) \\
&\quad \times \prod_{k=1}^{m-1} q_{i_{k-1} i_k} t_{\lfloor J_{k+1} \rfloor + 1} \exp\big(- q_{i_k} D(J_k, J_{k+1})\big) \\
&\quad \times q_{i_{m-1} i_m} \exp\big(- q_{i_k} D(J_m, T^*)\big),
\end{aligned}$$

Let us denote the number of transitions from state i to state j in k-th year as $n_{ij}(k)$ and the time spend by the system in state i in k-th year as $R_i(k)$,

$$R_i(k) = \sum_{l=0}^{m} \big(\min\{J_{l+1}, k+1\} - \max\{J_l, k\} \big)^+ \mathbf{1}_{\{i_l = i\}}.$$

Note that

$$\begin{aligned}
n_{ij} &= \sum_{k=1}^{m} n_{ij}(k), \\
R_i &= \sum_{k=1}^{m} R_i(k).
\end{aligned}$$

Thus, the maximum likelihood function can be rewritten into

$$L(\mathbf{Q}) = \prod_{k=1}^{m} \prod_{i=1}^{K} \prod_{\substack{j=1 \\ j \neq i}}^{K} (t_k q_{ij})^{n_{ij}(k)} \exp\big(- t_k q_{ij} R_i(k)\big),$$

The log–likelihood is then

$$l(\mathbf{Q}) = \sum_{k=1}^{m} \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \Big(n_{ij}(k) \log(t_k q_{ij}) - t_k q_{ij} R_i(k) \Big). \tag{3.4}$$

After differentiating we get

$$0 = \frac{\partial l(\mathbf{Q})}{\partial q_{ij}} = \sum_{k=1}^{m} \left(\frac{n_{ij}(k)}{q_{ij}} - t_k R_i(k) \right), \qquad i \neq j,$$

and after rewriting

$$\begin{aligned}
\hat{q}_{ij} &= \frac{\sum_{k=1}^{m} n_{ij}(k)}{\sum_{k=1}^{m} t_k R_i(k)} = \frac{n_{ij}}{\sum_{k=1}^{m} t_k R_i(k)}, \qquad i \neq j, \\
\hat{q}_{ii} &= - \sum_{j \neq i} \hat{q}_{ij}, \qquad i = 1, \ldots, K,
\end{aligned} \tag{3.5}$$

which is quite an intuitive result. In the numerator of (3.5) there is a number of transitions from state i to state j and in the denominator is a weighted sum of time spend in state i weighted by the time evolution. The classical ML estimator is the special case when $t_1 = t_2 = \ldots = t_n = 1$.

3.2.2 Time Evolution as the Model Parameter

In this case the time evolution is not given, but it is an additional set of parameters of the model, which needs to be estimated. Nevertheless, it is easy to extend the result from the previous paragraph.

From (3.4) we know that the log–likelihood is

$$l(\mathbf{Q}, \boldsymbol{t}) = \sum_{k=1}^{m} \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \Big(n_{ij}(k) \log(t_k q_{ij}) - t_k q_{ij} R_i(k) \Big).$$

Now let us differentiate it with respect to q_{ij} and t_k

$$0 = \frac{\partial l(\mathbf{Q}, \boldsymbol{t})}{\partial q_{ij}} = \sum_{k=1}^{m} \left(\frac{n_{ij}(k)}{q_{ij}} - t_k R_i(k) \right), \qquad i \neq j, \tag{3.6}$$

$$0 = \frac{\partial l(\mathbf{Q}, \boldsymbol{t})}{\partial t_k} = \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \left(\frac{n_{ij}(k)}{t_k} - q_{ij} R_i(k) \right). \tag{3.7}$$

From (3.6) we get the same to the previous section

$$q_{ij} = \frac{\sum_{k=1}^{m} n_{ij}(k)}{\sum_{k=1}^{m} t_k R_i(k)}, \qquad i \neq j. \tag{3.8}$$

From (3.7) we get for all $k \in \{1, \dots, m\}$

$$\sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \frac{n_{ij}(k)}{t_k R_i(k)} = \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} q_{ij}.$$

Note that the right side is the same for all k, hence it holds

$$\frac{1}{t_1} \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \frac{n_{ij}(1)}{R_i(1)} = \ldots = \frac{1}{t_m} \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \frac{n_{ij}(m)}{R_i(m)}.$$

Let $l \in \{1, \dots, m\}$ is arbitrary fixed. Then we can express every t_k using t_l

$$t_k = t_l \frac{\sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \frac{n_{ij}(k)}{R_i(k)}}{\sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \frac{n_{ij}(l)}{R_i(l)}}.$$

Now using the condition (3.3) we have

$$m = \sum_{k=1}^{m} t_k = t_l \frac{\sum_{k=1}^{m} \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \frac{n_{ij}(k)}{R_i(k)}}{\sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \frac{n_{ij}(l)}{R_i(l)}},$$

and hence

$$t_l = \frac{m \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \frac{n_{ij}(l)}{R_i(l)}}{\sum_{k=1}^{m} \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \frac{n_{ij}(k)}{R_i(k)}}.$$

Therefore, ML estimators in TEPG settings are

$$\hat{t}_l = \frac{m \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j\neq i}}^{K} \frac{n_{ij}(l)}{R_i(l)}}{\sum_{k=1}^{m} \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j\neq i}}^{K} \frac{n_{ij}(k)}{R_i(k)}}, \qquad l = 1, \ldots, m,$$

$$\hat{q}_{ij} = \frac{\sum_{k=1}^{m} n_{ij}(k)}{\sum_{k=1}^{m} \hat{t}_k R_i(k)}, \qquad i \neq j,$$

$$\hat{q}_{ii} = -\sum_{\substack{j=1 \\ j\neq i}}^{K} \hat{q}_{ij}, \qquad i = 1, \ldots, K.$$

3.2.3 Time Evolution as Random Variables

In the last case we assume that the time evolution is an non–decreasing process. We assume that for every $t > 0$ the time evolution T_t is a random variable with distribution $\mu(\boldsymbol{\alpha_t})$, which depends on a vector of parameters $\boldsymbol{\alpha_t}$. Note that $\boldsymbol{\alpha_t}$ can be time dependent and hence not stationary. First we describe the behavior of an exponentially distributed random variable under stochastic time and later CTMC under stochastic time.

Let τ is a positive random variable, which, conditioned on realization of the time evolution, has an exponential distribution i.e.,

$$\mathsf{P}[\tau \leq t | T_t] = 1 - \mathrm{e}^{-T_t \lambda}.$$

The unconditioned probability distribution of τ is then

$$\mathsf{P}[\tau \leq t] = \mathsf{E}\left[\mathsf{P}[\tau \leq t | T_t]\right] = 1 - \mathsf{E}\left[\mathrm{e}^{-T_t \lambda}\right].$$

We see that the only thing which we need to know is a Laplace transform of the random variable T_t at point λ. Let us denote by $\nu(t)$ the Laplace exponent of T_t at point λ and work with it as a function of time t,

$$\mathsf{E}\left[\mathrm{e}^{-T_t \lambda}\right] = \mathrm{e}^{-\nu(t)}.$$

Then

$$\mathsf{P}[\tau \leq t] = 1 - \mathrm{e}^{-\nu(t)}. \tag{3.9}$$

By differentiating (3.9) we get the density function of τ

$$f_\tau(t) = \mathrm{e}^{-\nu(t)} \nu'(t),$$

where $\nu'(t)$ is the derivative of the Laplace exponent with respect to time t. Note that we started our derivation at time 0. More general case would be if we started at some time $t_0 \geq 0$. Then we are interested in

$$\mathsf{P}[\tau \leq t | \mathcal{F}_{t_0}], \qquad t \geq t_0.$$

Since we allow general (dependent and non–stationary) time evolution increments, random time increment over interval $[t_0, t]$ can depend on σ-algebra $\mathcal{F}_{t_0}$. The dependency is through parameters $\boldsymbol{\alpha}_{t_0,t}$ of the time increment distribution, which we can think of as a functions of

$\mathcal{F}_{t_0}$ and t and denote it by $\boldsymbol{\alpha}(\mathcal{F}_{t_0}, t)$. Hence ν is a function of t, t_0, λ and $\mathcal{F}_{t_0}$. Then we have the density function of random time τ

$$f_\tau(t) = \mathrm{e}^{-\nu(t,t_0,\lambda,\mathcal{F}_{t_0})}\nu'(t, t_0, \lambda, \mathcal{F}_{t_0}), \qquad \text{for } t \geq t_0.$$

Now we can study CTMC under the random time evolution. Recall that the maximum likelihood function of the time–unchanged CTMC with observations up to horizon T^* (assuming we are given the initial state i_0) is

$$\begin{aligned} L(\mathbf{Q}) = q_{i_0} &\exp(-q_{i_0}J_1) \\ &\times \prod_{k=1}^{m-1} \frac{q_{i_{k-1}i_k}}{q_{i_{k-1}}} q_{i_k} \exp\big(-q_{i_k}(J_k - J_{k-1})\big) \\ &\times \frac{q_{i_{m-1}i_m}}{q_{i_{m-1}}} \exp\big(-q_{i_m}(T^* - J_m)\big). \end{aligned}$$

By time changing maximum likelihood changes into

$$\begin{aligned} L(\mathbf{Q}, \boldsymbol{\alpha}) = \nu'(J_1, 0, q_{i_0}, \mathcal{F}_0) &\exp\big(-\nu(J_1, 0, q_{i_0}, \mathcal{F}_0)\big) \\ &\times \prod_{k=1}^{m-1} \frac{q_{i_{k-1}i_k}}{q_{i_{k-1}}} \nu'(J_k, J_{k-1}, q_{i_k}, \mathcal{F}_{J_{k-1}}) \exp\big(-\nu(J_k, J_{k-1}, q_{i_k}, \mathcal{F}_{J_{k-1}})\big) \\ &\times \frac{q_{i_{m-1}i_m}}{q_{i_{m-1}}} \exp\big(-\nu(T^*, J_m, q_{i_m}, \mathcal{F}_{J_m})\big). \end{aligned} \tag{3.10}$$

For any further computation we need a particular choice of the time evolution distribution and mainly the dependency structure of parameters $\boldsymbol{\alpha}(\mathcal{F}_{t_0}, t)$, because it is very inconvenient to work with σ-algebra as a variable. Some feasible dependency structure could be a case when time evolution would be a Markov process. Then

$$\boldsymbol{\alpha}(\mathcal{F}_{t_0}, t) = \boldsymbol{\alpha}(T_{t_0}, t).$$

Even then the computation could be very difficult. We can get more explicit results if T_t is a process with independent stationary increments, i. e., the Lévy subordinator.

Case of Lévy Subordinator

Let us compute the maximum likelihood estimator explicitly for the case when time evolution is modeled through an α-stable subordinator. We have to estimate parameters α and q_{ij}. Recall that we need to add an additional condition to get unique estimators. We can not use condition (3.3) since time evolution is random and not parameters. We need a condition on parameters which are here generator matrix $\mathbf{Q}$ and parameter α. Let $\alpha = c$, where $0 < c \leq 1$. The particular choice of constant c is discussed later. Recall that the Laplace exponent ν does not depend on t_0 and parameters of time evolution does not depend on $\mathcal{F}_{t_0}$, but depends on constant α. We have

$$\begin{aligned} \nu(t, \lambda, \alpha) &= t\lambda^\alpha, \\ \nu'(t, \lambda, \alpha) &= \lambda^\alpha. \end{aligned}$$

By substitution in (3.10) we get

$$\begin{aligned}
L(\mathbf{Q},\alpha) &= q_{i_0}^{\alpha}\exp(-J_1 q_{i_0}^{\alpha}) \\
&\quad\times\prod_{k=1}^{m-1}\frac{q_{i_{k-1}i_k}}{q_{i_{k-1}}}q_{i_k}^{\alpha}\exp\big(-(J_k-J_{k-1})q_{i_k}^{\alpha}\big) \\
&\quad\times\frac{q_{i_{m-1}i_m}}{q_{i_{m-1}}}\exp\big(-(T^*-J_m)q_{i_m}^{\alpha}\big) \\
&= q_{i_0}^{\alpha-1}\exp(-J_1 q_{i_0}^{\alpha}) \\
&\quad\times\prod_{k=1}^{m-1}q_{i_{k-1}i_k}q_{i_k}^{\alpha-1}\exp\big(-(J_k-J_{k-1})q_{i_k}^{\alpha}\big) \\
&\quad\times q_{i_{m-1}i_m}\exp\big(-(T^*-J_m)q_{i_m}^{\alpha}\big) \\
&= \prod_{i=1}^{K}\exp\big(-q_i^{\alpha}R_i\big)\prod_{\substack{j=1\\ j\neq i}}^{K}\big(q_{ij}q_j^{\alpha-1}\big)^{n_{ij}}.
\end{aligned}$$

The log–likelihood is then

$$\begin{aligned}
l(\mathbf{Q},\alpha) &= \sum_{i=1}^{K}\Big(-q_i^{\alpha}R_i+\sum_{\substack{j=1\\ j\neq i}}^{K}n_{ij}\big(\log(q_{ij})+(\alpha-1)\log(q_j)\big)\Big) \\
&= -\sum_{i=1}^{K}\Big(\sum_{\substack{j=1\\ j\neq i}}^{K}q_{ij}\Big)^{\alpha}R_i+\sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K}n_{ij}\log(q_{ij}) \\
&\quad+(\alpha-1)\sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K}n_{ij}\log\Big(\sum_{\substack{k=1\\ k\neq j}}^{K}q_{jk}\Big),
\end{aligned}$$

where we used that $q_i=\sum_{\substack{j=1\\ j\neq i}}^{K}q_{ij}$. This relation we will use also the other way around.

If we differentiate the log–likelihood with respect to q_{ij} and set the result equal to 0 we get equations

$$\frac{\partial l(\mathbf{Q},\alpha)}{\partial q_{ij}}=-\alpha q_i^{\alpha-1}R_i+\frac{n_{ij}}{q_{ij}}+\frac{(\alpha-1)}{q_i}\sum_{\substack{k=1\\ k\neq i}}^{K}n_{ki}=0,\qquad i\neq j,$$

which we can rewrite using notation

$$n_{i.}=\sum_{\substack{j=1\\ j\neq i}}^{K}n_{ij},\qquad n_{.i}=\sum_{\substack{j=1\\ j\neq i}}^{K}n_{ji},$$

into the form

$$\alpha q_i^{\alpha-1}R_i-\frac{(\alpha-1)}{q_i}n_{.i}=\frac{n_{ij}}{q_{ij}}. \tag{3.11}$$

From (3.11) we see that for fixed i the ratio $\frac{n_{ij}}{q_{ij}}$ is constant for every j. Let $m \in \{1, 2, \ldots, i-1, , i+1, \ldots, K\}$ be an arbitrary index different than i. Then it holds that

$$q_{ij} = \frac{n_{ij}}{n_{im}} q_{im}, \qquad j \neq i. \tag{3.12}$$

If we substitute (3.12) into Equation (3.11) associated with indices i and m and rewrite it a bit, we get

$$\begin{aligned}
\alpha q_i^{\alpha-1} R_i - \frac{(\alpha-1)}{q_i} n_{\cdot i} &= \frac{n_{im}}{q_{im}}, \\
\alpha \Big(\sum_{\substack{j=1 \\ j \neq i}}^{K} q_{ij} \Big)^{\alpha-1} R_i - \frac{(\alpha-1)}{\sum_{\substack{j=1 \\ j \neq i}}^{K} q_{ij}} n_{\cdot i} &= \frac{n_{im}}{q_{im}}, \\
\alpha \Big(\sum_{\substack{j=1 \\ j \neq i}}^{K} \frac{n_{ij}}{n_{im}} q_{im} \Big)^{\alpha-1} R_i - \frac{(\alpha-1)}{\sum_{\substack{j=1 \\ j \neq i}}^{K} \frac{n_{ij}}{n_{im}} q_{im}} n_{\cdot i} &= \frac{n_{im}}{q_{im}}, \\
\alpha \Big(\frac{q_{im}}{n_{im}} \Big)^{\alpha} n_{i\cdot}^{\alpha-1} R_i - (\alpha-1) \frac{n_{\cdot i}}{n_{i\cdot}} &= 1.
\end{aligned}$$

After few more manipulations we get

$$\begin{aligned}
q_{im}^{\alpha} &= \frac{n_{im}^{\alpha} \big(n_{i\cdot} - (1-\alpha) n_{\cdot i} \big)}{\alpha R_i n_{i\cdot}^{\alpha}}, \\
q_{im} &= \frac{n_{im}}{n_{i\cdot}} \sqrt[\alpha]{\frac{n_{i\cdot} - (1-\alpha) n_{\cdot i}}{\alpha R_i}}.
\end{aligned} \tag{3.13}$$

There is one thing which we need to take care of. One could argue that under the root there might be a negative number, which depends on our choice of $\alpha = c$. However, note that under the root, there is a difference between number of how many times system entered (multiplied by $1-c$) and exited the state i. The difference $n_{i\cdot} - n_{\cdot i}$ can not be smaller than -1, since every time when the process visited (entered and exited) the state i both $n_{i\cdot}, n_{\cdot i}$ increased. The negative difference can occur only when system started in a different state than i and finished in i. In that case $n_{i\cdot} + 1 = n_{\cdot i}$ and we have under the root

$$\begin{aligned}
n_{i\cdot} - (1-\alpha) n_{\cdot i} &\geq 0 \\
n_{i\cdot} - (1-\alpha)(n_{i\cdot} + 1) &\geq 0 \\
\alpha &\geq \frac{1}{n_{i\cdot} + 1}.
\end{aligned}$$

Since we assume that $n_{im} > 0$ (otherwise $q_{im} = 0$),

$$\frac{1}{n_{i\cdot} + 1} \leq \frac{1}{2}.$$

Hence we have the ML estimator

$$\hat{\alpha} = c, \qquad c \in \left[\frac{1}{2}, 1\right]$$
$$\hat{q}_{im} = \frac{n_{im}}{n_{i.}} \sqrt[c]{\frac{n_{i.} - (1-c)n_{.i}}{cR_i}}, \qquad i \neq m,$$
$$\hat{q}_{ii} = -\sum_{\substack{m=1 \\ m \neq i}}^{K} \hat{q}_{im}, \qquad i = 1, \ldots, K.$$

Note that for $c = 1$ the Lévy subordinator degenerate to the deterministic identical function and we get the classical maximum likelihood estimator for the continuous–time Markov chain

$$\hat{q}_{im} = \frac{n_{im}}{R_i}.$$

An open question (for further research) remains the optimal choice of c and how the particular choice influences the performance of the ML estimator.

We derived ML estimators under three possible approaches to the time evolution. Nevertheless, for ML estimation we need completely observed data, which is usually not the case in practice, where system is observed in a regular period. Under this partial observation ML estimator is not applicable. Though, it is often applied and hence a systematic error is introduced.

3.3 Partially Observed Data under Stochastic Time

In this section we will try to extend methods from Section 2.3 in a way of Section 3.1.2. Recall that all six methods from Section 3.1.2 assume that we are given a one–year partial observation of the system. Methods DA, WA, QO and CO estimate (using the cohort method) matrix $\hat{\mathbf{P}}$ and then look for a possible generator $\hat{\mathbf{Q}}$ which generate a transition probability matrix as close as possible to $\hat{\mathbf{P}}$. The EM and MCMC methods rather estimate the generator directly. Let us try to extend this problem to a multi–year setting. We do not describe the whole algorithms again, but rather only parts which differ from one–year settings from Section 2.3.

We assume that we are given the discrete yearly observations from a sequence of m years. Further, we assume that the system follows the time–changed CTMC.

Let $\mathbf{Q}$ be the valid generator matrix and $T_1, \ldots, T_m$ be some non–negative random variables (yearly increments of time evolution process). We do not require them to be independent or identically distributed. Let $t_1, \ldots, t_m$ be their realizations. Then we assume, that in the i-th year, the system follows (conditioned on time evolution realization t_i) the continuous–time Markov chain with the generator $t_i\mathbf{Q}$.

We would like to estimate unknown parameters. Same to the Section 3.2 we distinguish cases when the time evolution is exogenously given (TEEG), is a parameter of the model (TEPG) or random variables (TERV). For each method we discuss all three approaches. However, as we see later, not all three approaches are always possible to apply. For DA method and WA method neither of approaches is applicable.

3.3.1 Quasi–optimization Method

The advantage which makes QO methods useful is that it allows us to split the optimization problem into subproblems on the level of rows. If we add the unknown time evolution into the

model, this advantage will disappear, since the split into subproblems is no longer possible. Hence, the only settings, which make sense, is TEEG, where we know $t_1, \ldots, t_m$ and we solve problem

$$\arg\min_{\mathbf{Q}\in\mathcal{Q}} \left(\sum_{i=1}^{m} \left\| t_i \mathbf{Q} - \log \hat{\mathbf{P}}_i \right\| \right) = \arg\min_{\mathbf{Q}\in\mathcal{Q}} \left(\sum_{i=1}^{m} \left\| \mathbf{Q} - \frac{1}{t_i} \log \hat{\mathbf{P}}_i \right\| \right),$$

which looks very similar to the one–year setting. There is just additional sum over years. And that sum is the biggest problem. QO algorithm is based on two lemmas from Tuenter (2001), which does not hold here. Hence neither TEEG settings is directly applicable, unless some alternative of these 2 lemmas are proven to hold and algorithm is extended. It could be another direction of further research.

3.3.2 Expectation Maximization Method

The EM method starts with some initial generator $\mathbf{Q}_0$ and then the EM algorithm iterates two steps. In the first step the expected values of n_{ij} and R_i are computed conditioned on the fact that generator is equal to $\mathbf{Q}_{n-1}$ (generator from previous iteration). In the second step ML estimator of $\mathbf{Q}_n$ is computed, where instead of n_{ij} and R_i are used conditioned expected values of n_{ij} and R_i computed in the first step. These two steps are iterated until convergence is reached. The extension into time change is very simple, since in Section 3.2 we derived all necessary ML estimators and all estimators are always functions of n_{ij} and R_i (or $n_{ij}(k)$ and $R_i(k)$). For each settings we describe only these two steps. Main iterative algorithm remains untouched. We denote the generator matrix in k-th year as $\mathbf{Q}(k) = t_k \mathbf{Q}$.

TEEG

When we are given by the time evolution, the first step is trivial. In the n-th iteration we set for each year $k = 1, \ldots, m$ $\mathbf{Q}_{n-1}(k) = t_k \mathbf{Q}_n$, and for each year we compute $\mathsf{E}\left[n_{ij}(k) | \boldsymbol{x}^{obs}, \mathbf{Q}_{n-1}(k)\right]$ and $\mathsf{E}\left[R_i(k) | \boldsymbol{x}^{obs}, \mathbf{Q}_{n-1}(k)\right]$ using (2.14) - (2.18).[19] In the second step of the n-th iteration we apply the ML estimator from Section 3.2.1.

$$q_{ij} = \frac{\sum_{k=1}^{m} n_{ij}(k)}{\sum_{k=1}^{m} t_k R_i(k)},$$

and get $\mathbf{Q}_n$.

TEPG

We do exactly same thing as in TEEG except in the second step we apply also estimator for time evolution from section 3.2.2.

TERV

Also in the random setting we do the similar thing as in the previous two cases, only in the second step we apply estimators from section 3.2.3, which depends on the choice of the time evolution distribution (In Section 3.2.3 we derived explicit estimators for α-stable subordinator).

[19] $\boldsymbol{x}^{obs}$ are observed data, see Section 2.3.3.

3.3.3 Markov Chain Monte Carlo Method

Recall that in the one–year MCMC method we used Beyesian settings

$$\begin{aligned} p(\mathbf{Q}|X, \boldsymbol{x}^{obs}) &\propto p(\mathbf{Q})p(X|\mathbf{Q}) \\ &\propto p(\mathbf{Q})L(\mathbf{Q}). \end{aligned}$$

where the posterior probability $p(\mathbf{Q}|X, \boldsymbol{x}^{obs})$ is up to a constant equal to the product of the prior probability $p(\mathbf{Q})$ and the likelihood function $L(\mathbf{Q})$. For easy use of this property we need the right choice of the prior distribution for matrix $\mathbf{Q}$. In the classical usage the obvious choice was a gamma distribution. In the time–changed settings we would like to have also the situation when the product of the prior distribution and the maximum likelihood will give us again a distribution from the same distribution family as the prior distribution is from. Further recall that in every iteration of the algorithm we update parameters of the prior distribution. In the classical MCMC (see Section 2.3.4) we have the prior distribution of $\mathbf{Q}$

$$p(\mathbf{Q}) \propto \prod_{i=1}^{K} \prod_{\substack{j=1 \\ j\neq i}}^{K} q_{ij}^{\alpha_{ij}-1} \exp(-q_{ij}\beta_i),$$

the maximum likelihood function

$$L(\mathbf{Q}) = \prod_{i=1}^{K} \prod_{\substack{j=1 \\ j\neq i}}^{K} q_{ij}^{n_{ij}} \exp(-q_{ij}R_i),$$

and the parameter update

$$\begin{aligned} \alpha_{ij} &= \alpha_{ij} + n_{ij}, \\ \beta_i &= \beta_i + R_i. \end{aligned}$$

TEEG

For TEEG we have the maximum likelihood function (see Section 3.2.1)

$$\begin{aligned} L(\mathbf{Q}) &= \prod_{k=1}^{m} \prod_{i=1}^{K} \prod_{\substack{j=1 \\ j\neq i}}^{K} (t_k q_{ij})^{n_{ij}(k)} \exp\big(-t_k q_{ij} R_i(k)\big) \\ &= \prod_{i=1}^{K} \prod_{\substack{j=1 \\ j\neq i}}^{K} \left(\prod_{k=1}^{m} t_k^{n_{ij}(k)} \right) q_{ij}^{n_{ij}} \exp\left(-q_{ij} \sum_{k=1}^{m} t_k R_i(k) \right), \end{aligned}$$

and since $t_1, \ldots, t_m$ are given we see that good choice of the prior distribution of q_{ij} is again the gamma distribution. In every iteration the parameter update is then

$$\begin{aligned} \alpha_{ij} &= \alpha_{ij} + n_{ij}, \\ \beta_i &= \beta_i + \sum_{k=1}^{m} t_k R_i(k). \end{aligned}$$

TEPG and TERV

Situation seems more complicated in the TEPG and TERV settings. In the TERV setting, situation highly depends on the choice of time evolution distribution. In both cases we would like to choose the prior distribution in the way that it can be easily updated by the maximum likelihood function. In the TEPG setting it means that the prior joint distribution of $\mathbf{Q}$ and $\boldsymbol{t}$ is up to a constant equal to

$$f(\mathbf{Q}, \boldsymbol{t}) = \prod_{k=1}^{m} \prod_{i=1}^{K} \prod_{\substack{j=1 \\ j \neq i}}^{K} (t_k q_{ij})^{n_{ij}(k)} \exp\big(- t_k q_{ij} R_i(k)\big).$$

In the TERV setting we would get the prior distribution in a similar form dependent on the time evolution distribution. Now the main problem is how to generate random variables with this joint density function. In the TEEG setting we could separate problem and generate every $q_{ij}, i \neq j$ separately. Here it is not possible. We need to generate $\mathbf{Q}$ and $\boldsymbol{t}$ together which can be done for example using the rejection method (see Devroye (1986, Section 2.3)). The parameters update then proceed in a similar way as in the case of TEEG.

3.3.4 Componentwise Optimization Method

The last method is the CO method. This method tries to solve the optimization problem by performing optimization in each parameter separately with other parameters fixed instead of the optimization over all parameters at the same time. The disadvantage of the method is that we don't have sure convergence of the method and even if it converges, it converges only to some local minimum, not necessary to the global minimum. However, if the convergence is reached the solution given by this method can not be worse than the starting point. If we use as a starting point solution of any previous methods we will get only better result. Better in the sense that the transition probability matrix given by this generator will be in Euclidean norm closer to the empirical transition probability matrix.

There is no way how to perform the componentwise optimization method in TERV setting. The only possible way is to find estimators of the time evolution in TEPG setting and after that try to model these estimators by another method (ML, moment method,...).

The algorithm in TEEG and TEPG setting almost does not differ. The TEPG settings have only one additional step. Hence we describe both settings together.

Let us recall that we have m empirical transition matrices $\hat{\mathbf{P}}_1, \dots, \hat{\mathbf{P}}_m$ and we want to find the generator matrix $\hat{\mathbf{Q}}$ and vector $\hat{\boldsymbol{t}}$, such that the transition probability matrix $\exp(\hat{t}_i \hat{\mathbf{Q}})$ is as near as possible to $\hat{\mathbf{P}}_i$ for $i = 1, \dots, m$. We are looking for arguments of

$$\min \left\| \sum_{k=1}^{m} \big(\exp(t_k \hat{\mathbf{Q}}) - \hat{\mathbf{P}}_k \big) \right\|,$$

over all possible values ($\mathbf{Q}$ needs to be the valid generator and t_k positive). In the TEEG settings t_k is given and we optimize only over possible values of $\mathbf{Q}$. Theoretically possible values of t_k or q_{ij} are unbounded and can be any positive real number. For the practical computation and application of optimization algorithm we need to bound the possible values (same as in one year settings). These boundaries should be rather conservative and can be set from the past observation and expectation. The higher (and hence safer) the bound will be, the slower the algorithm will proceed. Let us denote the bound for parameter q_{ij} by c and for t_k by d.

Now we need a starting point for matrix $\mathbf{Q}$. One possibility is to take the result of any previous method. In the TEPG settings we need to get time evolution estimation. For that, we perform the following optimization

- For $k = 1, \ldots, m$ do
  - Solve the one–dimensional optimization problem

$$\hat{t}_k = \arg \min_{0 \le t \le d} \left\| \exp(t\hat{\mathbf{Q}}) - \hat{\mathbf{P}}_k \right\|.$$

- Rescale vector $\boldsymbol{t}$ for $\sum_{k=1}^{m} t_k = m$ to hold

$$D = \sum_{i=1}^{m} \hat{t}_i,$$

and then set

$$\boldsymbol{t} = \hat{\boldsymbol{t}} \frac{m}{D},$$
$$\mathbf{Q} = \hat{\mathbf{Q}} \frac{D}{m}.$$

The CO proceeds in the following way:

1. Set starting point $\mathbf{Q}$.
2. Until the convergence is reached, repeat
   - For $i = 1, \ldots, K$ do
     - For $j = 1, \ldots, K$ do
       - If $i \neq j$, find q_{ij} as

$$q_{ij} = \arg \min_{0 \le q_{ij} \le c} \sum_{k=1}^{m} \left\| \exp\left(t_k \mathbf{Q}(q_{ij})\right) - \hat{\mathbf{P}}_k \right\|.$$

         Simultaneously with changing q_{ij} adjust the other row elements of $\mathbf{Q}$ to remain a valid generator.
   - In TEPG settings perform the time optimization and time rescaling as described above
3. Return $\mathbf{Q}$ as the optimal solution.

4
Credit Risk Models

First, we are going to introduce some basic concepts and ideas of credit risk modeling. Then we will show what are commonly used practices in modeling the loss distribution.

4.1 Idea of Credit Risk

One of the main businesses of banks is lending money. They lend money to people or firms that need money which they do not have. Of course banks lend money just to debtors that have a high probability to repay it. Ability of repaying money is called the *credit worthiness*. For lending money, the bank expects some rewards in the form of interest. The interest is composed of reward for lending and some risk premium for bearing the risk that the debtor will fail to satisfy his obligation and a loss for the bank will occur.

Usually a contract develops in the following way. A future debtor applies for a loan. The bank checks his credit worthiness. It consists of checking the present financial situation, evaluation of present and estimation of future assets and liabilities etc.[20] If the debtor fulfills all internal requirements of the bank loan, there are couple of things that need to be established — the height of the loan, the height of the interest rate, the calendar of instalment, the conditions of drawing money, etc. After signing the contract the debtor withdraws the money and starts paying regular instalments according to the calendar of instalments until the whole debt is repayed including the interest.

That is the optimal development of the contract, but sometimes deviations from the optimal development might occur. The debtor can get into financial distress and fail to fulfill his obligation. Usually at the time when the debt is granted the debtor has a high credit worthiness (is in a good financial condition), but during the duration of the contract the credit worthiness can change. The change in credit worthiness is often called a *credit event*. A situation when the debtor fails to fulfill his obligation is called a *default* and the random time at which it happens is called the *default time*.

A default of the debtor means a financial loss for the bank. This loss is called the *credit loss* and we will denote it by L. When a default occurs it does not mean that all outstanding money is lost, some fraction can be recovered. The fraction that is recovered is called the *recovery rate* (RR) and it is from the interval $[0, 1]$. The height of the credit loss, if a default occurs, is a random variable which depends on many factors, mainly on the *exposure at default* (EAD) and the *loss given default* (LGD). The EAD is the outstanding amount of drawn money at the time of default. The LGD is simply $1 - RR$, i.e., the fraction of the EAD that will be lost at default. The LGD depends on many circumstances like the presence of a collateral, the business sector of the debtor, the balance sheet of the debtor, etc.

Some debts have a valuable collateral which covers a big portion of the debt. Then it appears to be useful to estimate the total amount C, which we can recover from the collateral

[20] The most commonly used tool for that are so called scoring cards, which are models based on logistic regression or a similar theory. Once all necessary details are given to model, the model gives a score for client. If the score is sufficiently high, the loan is granted. If it is to low, the loan is rejected, and if it is in the middle further personal examination is performed.

for sure and then define a new $\widetilde{EAD}$ as

$$\widetilde{EAD} = (EAD - C)^+.$$

Then we assume just the new $\widetilde{EAD}$ in the model. Sometimes a collateral value is very volatile. In that case the estimate of the amount C should be very conservative even if we know that with the very high probability we will be able to recover a bigger amount than C at default.

The bank has a whole portfolio of loans denoted by $PF = \{1, \ldots, N\}$. One debtor can have more loans but we can view it as one loan with instalments that came from more loans. The credit loss L^i on loan $i \in PF$ can be written in following way

$$L^i = EAD_i \times LGD_i \times \mathbf{1}_{\{\tau_i \leq T_i\}},$$

where τ_i is the default time of the debtor (possibly infinite if the debtor will not default) and T_i is the maturity of loan i. The credit loss of the whole portfolio is simply a sum of losses over all loans. Thus,

$$L = \sum_{i \in PF} EAD_i \times LGD_i \times \mathbf{1}_{\{\tau_i \leq T_i\}}.$$

The credit portfolio loss L is also called the *aggregate loss*. A portfolio credit risk model is interested in the estimation of the probability distribution of L and the time evolution of the loss. Let us introduce the loss process of loan i:

$$L_t^i = EAD_i \times LGD_i \times \mathbf{1}_{\{\tau_i \leq \min\{t, T_i\}\}}. \tag{4.1}$$

Then the loss process of the whole portfolio is

$$L_t = \sum_{i \in PF} EAD_i \times LGD_i \times \mathbf{1}_{\{\tau_i \leq \min\{t, T_i\}\}}.$$

Note that the loss process L_t is a non–decreasing stochastic process. Since L_t is a stochastic process we can try to compute its expected value $\mathsf{E}\, L_t$ which we call the *expected loss*. The expected loss is something which a bank should count with in its balance sheet and future cash flow planning. We can also define the *unexpected loss* as a standard deviation of L_t. The unexpected loss shows how volatile the credit loss is and how much it probably will differ from its expected value $\mathsf{E}\, L_t$. The bank should, and as a matter of fact it is obliged, to have reserves for covering the unexpected loss.

It is useful to know the whole distribution of the process L_t since the bank managers are often interested in other statistics then the expected and unexpected loss. The main example is the credit VaR — *Value at Risk* which is a loss boundary which will not be crossed with a very high probability, typically 99%. More rigorously the credit VaR_t^α is defined as

$$\mathrm{VaR}_t^\alpha = \inf \left\{ l \in \mathbb{R} | \mathsf{P}[L_t > l] < 1 - \alpha \right\}.$$

The VaR is a very commonly used risk measure in practice, but it has a big disadvantage since it does not tell anything about the height of the loss if the loss is higher than VaR. It can be fixed by introducing the risk measure CVaR — *Conditional Value at Risk* that is also often called the *Expected Shortfall* and is given by

$$\mathrm{CVaR}_t^\alpha = \mathsf{E}\left[L_t | L_t > \mathrm{VaR}_t^\alpha \right].$$

There are many ways to model the process L_t. One can try to model it as an aggregate process which we will refer to as *aggregate models*. The idea is similar to one where we would like to model the movement of a stock index that consists of many stocks. One way is to model the movement of every stock and the correlations between them. Another way is to model the index as some stochastic process. In that case some information is lost, but the modeling is easier. The aggregate loss process can be modeled using many approaches — time series, non–decreasing Markov chains, (Lévy) subordinators. We are not going to describe these models in this thesis and we refer for an overview to Giesecke (2008), where also other references can be found.

Other models describe the evolution of every particular loan and the dependency between them. There are three main types of these models — *structural models* (Section 4.3), *reduced–form models* (Section 4.4), and incomplete information models (Section 4.5).

4.2 From Loan to Defaultable Zero–coupon Bond

Before we start describing portfolio credit risk models we should say a few words about the pricing of credit derivatives. We expect that the reader is familiar with basic ideas and techniques of risk neutral pricing. For an introduction to risk neutral pricing we refer to Baxter and Rennie (1996), Shreve (2004), Musiela and Rutkowski (2005), or any other financial stochastic introductionary text.

4.2.1 Zero–coupon Bonds

We assume an arbitrage free economy with the risk free interest rate r_t that can follow any short rate model as well as more complex HJM interest rate models. Then we can define a *saving account* B_t as

$$B_t = \exp\left(\int_0^t r_s \,\mathrm{d}s\right),$$

that expresses the time value of the money. If we have 1 unit of money at time 0, we will have without any risk B_t units of money at time t. Conversely 1 unit of money at time t is worth B_t^{-1} at time 0.

We know that an absence of the arbitrage opportunity is equivalent to the existence of the risk neutral probability measure P^* and also implies the uniqueness of the risk free interest rate r_t. For an overview of interest rate models we refer to Musiela and Rutkowski (2005) or Brigo and Mercurio (2006) where all models are with respect to Brownian motions. If we want to move to more general Lévy processes in an interest rate framework, see for example Kluge (2005).

From the theory of risk neutral pricing we know that an arbitrage free price $\pi_t(X)$ of the European contingent claim X settled at time T is given by pricing formula

$$\pi_t(X) = B_t \,\mathsf{E}_{\mathsf{P}^*}\left[B_T^{-1} X | \mathcal{F}_t\right]. \tag{4.2}$$

The zero–coupon bond is a security which has pay–off 1 at maturity T. Using formula (4.2) we can deduce that the arbitrage free price of the zero–coupon bond at time t is

$$B(t,T) = B_t \,\mathsf{E}_{\mathsf{P}^*}\left[B_T^{-1} | \mathcal{F}_t\right]. \tag{4.3}$$

Zero–coupon bonds are in fact very rarely traded in markets but they are very important modeling tools derived from coupon bonds or other interest rate securities. Bonds can be

issued by governments, banks or companies. One of the reasons why bonds are so important is because any fixed–income security can be written as some portfolio of zero–coupon bonds with different maturities $T_1, \dots, T_n$. A payoff of the security is then given by

$$S_t = \sum_{i=1}^{n} c_i \mathbf{1}_{\{t=T_i\}},$$

where c_i is the height of the payment at time T_i. We can think about every payment as c_i zero–coupon bonds with maturity T_i. Hence the arbitrage free price of the security S at time t is

$$\pi_t(S) = \sum_{i=1}^{n} c_i B(t, t_i).$$

Defaultable Zero–coupon Bond

As far as bonds are issued by some issuer there is always a risk that the issuer goes into default and will fail to pay the bond. For some issuers, such as the US government or big international banks, the default (or credit) risk is negligible and we call their bonds risk free bonds and formula (4.3) is valid for them. For the rest, we need to incorporate the default risk into the valuation formula. For the discussion above, it is enough, if we investigate just the pricing formula of the *defaultable zero–coupon bond*. Coupon bonds can be built from zero–coupon bonds. The defaultable zero–coupon bond has pay–off 1 if the default time τ of the issuer is higher than the maturity T and the recovery RR otherwise, i.e.,

$$\mathbf{1}_{\{\tau>T\}} + RR\, \mathbf{1}_{\{\tau\leq T\}}.$$

Hence, using (4.2), the arbitrage free price of the defaultable zero–coupon bond is

$$\begin{aligned} D(t,T) &= B_t\, \mathsf{E}_{\mathsf{P}^*}\, \left[B_T^{-1}(\mathbf{1}_{\{\tau>T\}} + RR\, \mathbf{1}_{\{\tau\leq T\}})|\mathcal{F}_t\right] \\ &= B_t\, \mathsf{E}_{\mathsf{P}^*}\, \left[B_T^{-1}(1-(1-RR)\mathbf{1}_{\{\tau\leq T\}})|\mathcal{F}_t\right] \\ &= B(t,T) - B_t\, \mathsf{E}_{\mathsf{P}^*}\, \left[B_T^{-1}(1-RR)\mathbf{1}_{\{\tau\leq T\}}|\mathcal{F}_t\right]. \end{aligned}$$

By comparing with (4.1) we see that the price of the defaultable zero–coupon bond is the price of the risk free zero–coupon bond minus a discounted expected loss computed with respect to the risk neutral measure P^*. The valuation for any fixed–income security can be established in the same manner as for risk free securities above.

Sometimes, bonds with the face value different than 1 will be mentioned. A bond with face value K is exactly the same as K standard bonds with face value 1.

From the no arbitrage condition we know that there exists a risk neutral probability measure P^*, but it is not unique in general. If we model an uncertainty in the model with a Brownian motion then market is complete, every contingent claim is replicable and the risk neutral measure is unique. But in a more general setting of Lévy processes there might exist infinitely many possible risk neutral measures (dependent if the market is or is not complete). Then we need to determine which measure we should use. A very widely used approach to choose a measure P^* is the Esscher transform.

If we fix t and T and compare the yield to maturity on the risk free zero–coupon bond and the defaultable zero–coupon bond, we can see that the defaultable zero–coupon bond is cheaper, hence if a default did not occur, the defaultable zero–coupon bond has a higher yield

to maturity. The difference between yields of the defaultable zero–coupon bond and the risk free zero–coupon bond is called a *credit spread* $S(t,T)$ and is equal to

$$S(t,T) = -\frac{\log D(t,T) - \log B(t,T)}{T-t}.$$

4.2.2 Loan as Defaultable Zero–coupon Bond

Some loans have a fixed interest rate and hence we know the whole instalment schedule. We can see these loans as fixed–income securities and model it as a portfolio of defaultable zero–coupon bonds, where every instalment is a zero–coupon bond.

Other loans have a floating interest rate. The floating interest rate is in the most cases the interest rate from the wholesale market (where the bank takes the money) plus the margin of the bank (incorporate the risk premium and margin for lending money). The future interest rate on the wholesale market can be estimated from forward rates which are observable on the market. The margin of the bank is constant in time in most cases. Hence we can estimate the interest rate on loans with a floating interest rate, from forward rates and then model it as a fixed–income security.

4.3 Structural Models

Structural models have roots in the early papers of Black and Scholes (1973) and Merton (1974). Structural models are very popular for their economic interpretation which are clearer than reduced–form modeling. The debt can be seen as a contingent claim on a firm's asset. The market value of a firm is the main source of uncertainty and a default occurs if the market value process V_t falls below some threshold. The biggest assumption and the main disadvantage of the model is that the firm's value process V_t has to be observable. An advantage of structural models over reduced–form models is that we do not need to specify recovery rates since they follow implicitly from the model as the residual value of the firm's asset at maturity.

A firm value process V_t is often modeled via a Brownian motion. The assumption of normality is very convenient to work with, but not very realistic and empirical studies show a significant deviation from normality. Continuous paths of a Brownian motion also imply a predictability of a default and vanishing credit spreads on bonds with near maturities. Empirical studies show that even on short term bonds there is a significantly high credit spread. A way to fix it, is to introduce jumps into a firm value process. For example Zhou (1997) assumed the Brownian part with an additional jump term. We will describe structural models in a more general framework of Lévy processes which involves a Brownian motion as a special case.

We assume that a firm value process follows a geometric Lévy process with respect to some real world measure P, i.e.,

$$V_t = V_0 \exp(X_t).$$

In all following models we are interested in expected values of future random variables. We could compute these expectations conditioned on information up to time $t \leq T$ (the σ-algebra $\mathcal{F}_t$), but since in our model's risk driver processes are Lévy processes with independent identically distributed increments it follows that for any integrable function g and $t \leq T$

$$\mathsf{E}\left[g(X_T - X_t)|\mathcal{F}_t\right] = \mathsf{E}\left[g(X_{T-t})\right].$$

Therefore we will assume that we are doing all computations for time 0, which allows for a simpler notation.

4.3.1 Merton's Model

The model introduced by Merton (1974) is interested in a firm value process only at the time of maturity T. Let us assume that a firm is financed by an equity and a debt with the profile of a zero coupon bond with face value K. A default occurs if the firm value process at the time of maturity is below the face value of the bond, hence the default time τ is defined as

$$\tau = \begin{cases} T, & V_T < K, \\ \infty, & V_T \geq K. \end{cases}$$

In case of a default, the equity is useless and the remaining value of assets goes to the creditor, otherwise the debt is repaid in full amount K and the amount $V_T - K$ belongs to the stockholders. Therefore, the pay-off of the defaultable zero–coupon bond at maturity is

$$\min(V_T, K) = K - (K - V_T, 0)^+,$$

and the pay–off of the equity is

$$(V_T - K)^+.$$

One can see that the bond's pay–off at maturity is the face value of the bond lowered by the pay–off of the put option on the firm's values with strike K and the pay–off of equity is the pay–off of the call option on the firm's value. This approach is often called the option theoretic approach or the firm value approach.

The probability of default is the probability that V_T will be below K,

$$DP = \mathsf{P}[V_T < K] = \mathsf{P}[V_0 \exp(X_T) < K] = \mathsf{P}\left[X_T < \log \frac{K}{V_0}\right],$$

which is equal to the cumulative distribution function F_{X_T} of X_T if F_{X_T} is continuous in the point $\log(K/V_0)$. Otherwise the left limit is chosen as its value.

The expected loss on the loan computed at time 0 is equal to the expected pay–off of the put option on the firm's value with respect to a real world probability measure P

$$\mathsf{E}\, L = \mathsf{E}\left[(K - V_T)^+\right].$$

Recall that we assume the dynamics $V_T = V_0 \exp(X_T)$. If the Lebesgue density f of X_T exists, the expected loss is

$$\mathsf{E}\, L = \int_{-\infty}^{\infty} (K - V_0 \mathrm{e}^x)^+ f(x)\, \mathrm{d}x,$$

and the expected return of the bond is $K - \mathsf{E}\, L$.

From a pricing point of view we are interested in a fair present value of the defaultable zero–coupon bond using risk–neutral pricing techniques. In a similar manner as in Section 4.2.1 we conclude that the value of the defaultable bond with face value K at time 0 is

$$D(0,T) = KB(0,T) - \mathsf{E}_{\mathsf{P}^*}\left[\frac{1}{B_T}(K - V_T)^+\right]. \tag{4.4}$$

There remain a couple of difficulties to solve. First of all we need to find a risk neutral measure P^*. Under the no arbitrage condition there exists exactly one risk neutral measure in a complete market. Unfortunately a complete market is rather an exception when we use a general Lévy process X_t. In an incomplete market there exist infinitely many risk neutral

measures. A widely–used approach to choose the risk neutral measure is the Esscher transform. Then we have a problem of the option pricing using the Esscher transform, see Gerber and Shiu (1994), Buhlmann et al. (1996), Miyahara (2001) and Elliott et al. (2005). Another problem comes up if we do not assume independence of B_T and V_T in (4.4). Then we need to compute their joint distribution, which is also not easy. In the case when the interest rate is a constant r and the risk neutral measure $\mathsf{P}^* = \mathsf{P}_\theta$ is chosen by the Esscher transform we have

$$\mathsf{E}_{\mathsf{P}^*}\left[\frac{1}{B_T}(K - V_T)^+\right] = \mathsf{E}\left[\frac{\exp(\theta X_T)}{\mathsf{E}\,\exp(\theta X_T)}\mathrm{e}^{-rT}\big(K - V_0\exp(X_T)\big)^+\right],$$

where θ is chosen such that the discounted firm value process V_t is a martingale with respect to the measure P_θ.

In the original paper of Merton (1974), the Lévy process X_t is assumed to be $X_t = (\mu - \sigma^2/2)t + \sigma W_t$, where $\mu \in \mathbb{R}$, $\sigma > 0$, and W_t is a standard Brownian motion. Since W_t is normally distributed with expected value 0 and variance t we have an equality in distribution

$$W_t \stackrel{d}{=} \sqrt{t}Y,$$

where Y is a standard normally distributed variable. In that setting the default probability DP is

$$\begin{aligned} DP &= \mathsf{P}\left[X_T < \log\frac{K}{V_0}\right] \\ &= \mathsf{P}\left[\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}Y < \log\frac{K}{V_0}\right] \\ &= \mathsf{P}\left[Y < \frac{\log\frac{K}{V_0} - \left(\mu - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}\right] \\ &= \Phi(d), \end{aligned}$$

where Φ is the cumulative distribution function of a standard normal distribution and

$$d = \frac{\log\frac{K}{V_0} - \left(\mu - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}.$$

The value d is often called the *distance to the default*. The expected loss is then

$$\begin{aligned} \mathsf{E}\,L &= \int_{-\infty}^{\infty}\left(K - V_0\mathrm{e}^{\left(\mu-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}x}\right)^+\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x \\ &= K\int_{-\infty}^{d}\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x \\ &\quad - V_0\int_{-\infty}^{d}\mathrm{e}^{\left(\mu-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}x}\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x \\ &= K\Phi(d) - V_0\mathrm{e}^{\mu T}\Phi\big(d - \sigma\sqrt{T}\big). \end{aligned} \tag{4.5}$$

Under a risk neutral measure P^* the firm value process evolves also as a geometric Brownian motion but with a different drift equal to the risk neutral interest rate, which is the result of Black and Scholes (1973). If we assume a constant interest rate r, the process X_t is

$$X_t = \left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t,$$

with respect to the measure P^*. Hence using (4.5), the expected loss of the bond with respect to measure P^* is

$$\mathsf{E}_{\mathsf{P}^*} L = K\Phi(d) - V_0 \mathrm{e}^{rT} \Phi\big(d - \sigma\sqrt{T}\big),$$

where

$$d = \frac{\log \frac{K}{V_0} - \left(r - \frac{\sigma^2}{2}\right) T}{\sigma\sqrt{T}}.$$

The price of the defaultable zero–coupon bond is the price of the risk free zero–coupon bond minus the discounted expected loss with respect to risk neutral measure P^*. Therefore

$$\begin{aligned} D(0,T) &= \mathrm{e}^{-rT} K - \mathrm{e}^{-rT} \left(K\Phi(d) - V_0 \mathrm{e}^{rT} \Phi\big(d - \sigma\sqrt{T}\big) \right) \\ &= \mathrm{e}^{-rT} K\Phi(-d) - V_0 \Phi\big(d - \sigma\sqrt{T}\big). \end{aligned}$$

The presented case where we have one firm that is financed by an equity and a bond is very simple. If we assume more firms we can easily compute the expected loss of the portfolio since the expected loss of portfolio is simply the sum of the expected losses of the particular loans. However, if we are interested in computing the unexpected loss, VAR or CVAR, of the portfolio we need to consider dependencies between the firm value processes. Also in practice, the financial structure of the firm is much more complicated than just an equity and bonds. In that case we have to take into account which liabilities of the firm have higher priority and include it in the model. For a discussion of a firm's liabilities see Vašíček (1984). Determining of the loss distribution is often done by simulation.

Extensions to Merton's Model

Many extensions to Merton's model have been done. For an overview of these extensions see Bielecki and Rutkowski (2002, Section 2.4) and references cited in there. One of the main extensions is presented in the next section.

4.3.2 First–Passage Model

In Merton's model a default can be observed only at maturity. Even if a firm's value almost vanishes a default is not triggered before maturity. To fix this, Black and Cox (1976) added a barrier D_t such that if value process V_t falls below it, a default is triggered and the creditor overtakes the control of the firm. The barrier D_t can be exogenously or endogenously given and can be constant, deterministic, or even random, but tractability rapidly decreases. The time when a firm value process V_t hits the barrier is called the *hitting time*. The hitting time probability distribution of a Lévy process is not known explicitly apart in a few exceptions.

In a first–passage model the default can be triggered in two ways. The first possibility is the same as in Merton's model at maturity T:

$$\tau_1 = \begin{cases} T, & V_T < K, \\ \infty, & V_T \geq K. \end{cases}$$

The second possible default is by hitting the barrier D_t

$$\tau_2 = \inf\left\{ t \middle| \, 0 < t < T, V_t < D_t \right\},$$

where we assume that the infimum of an empty set is equal to $+\infty$. Let us define a default for the first–passage model as

$$\tau = \min\{\tau_1, \tau_2\}.$$

We can also choose $D_T = K$ to default τ_1 and τ_2 agree at maturity T.

Let us assume a barrier as the face value of the bond discounted using some constant discount factor $\gamma > 0$, i.e.,

$$D_t = \mathrm{e}^{-\gamma(T-t)}K.$$

Then we have

$$\begin{aligned}
\mathsf{P}[V_t \le D_t] &= \mathsf{P}[V_0\mathrm{e}^{X_t} \le \mathrm{e}^{-\gamma(T-t)}K] \\
&= \mathsf{P}[V_0\mathrm{e}^{X_t-\gamma t} \le \mathrm{e}^{-\gamma T}K] \\
&= \mathsf{P}\left[X_t - \gamma t \le \log\left(\mathrm{e}^{-\gamma T}\frac{K}{V_0}\right)\right] \\
&= \mathsf{P}\left[\tilde{X}_t \le \log\tilde{K}\right],
\end{aligned}$$

where

$$\begin{aligned}
\tilde{X}_t &= X_t - \gamma t, \\
\tilde{K} &= \mathrm{e}^{-\gamma T}\frac{K}{V_0}.
\end{aligned}$$

We see that the probability distribution of hitting the barrier D_t of the original process V_t is the same as the probability distribution of hitting the constant barrier $\log\tilde{K}$ with the Lévy process $\tilde{X}_t$. Let us define the process $\tilde{M}_t$ as the running minimum of the process $\tilde{X}_t$

$$\tilde{M}_t = \min_{0\le s\le t}\tilde{X}_s.$$

A default occurs if and only if there exists $t \le T$ such that $V_t < D_t$ and it is equivalent with $\tilde{M}_t < \log\tilde{K}$. Hence the default probability is

$$DP = \mathsf{P}[\tilde{M}_T < \log\tilde{K}]. \tag{4.6}$$

If a default occurs the creditor takes control over the firm and his pay–off at maturity is V_T, otherwise the face value K is payed. Therefore the pay–off of the bond at maturity is

$$\begin{aligned}
&K\mathbf{1}_{\{\tilde{M}_T\ge\log\tilde{K}\}} + V_T\mathbf{1}_{\{\tilde{M}_T<\log\tilde{K}\}} \\
&\qquad = K - (K-V_T)^+ + (V_T-K)^+\mathbf{1}_{\{\tilde{M}_T<\log\tilde{K}\}} \\
&\qquad = K - (K-V_T)^+ + \mathrm{e}^{\gamma T}V_0\left(\mathrm{e}^{\tilde{X}_T} - \tilde{K}\right)^+\mathbf{1}_{\{\tilde{M}_T<\log\tilde{K}\}}.
\end{aligned} \tag{4.7}$$

One can see from (4.7) that the pay–off is at least as high as in Merton's model. Furthermore, the pay–off can be even higher than the face value of the bond and an additional revenue $(V_T - K)^+$ can occur. From (4.7) it follows that the expected loss in the model is

$$\mathsf{E}\,L = \mathsf{E}\,(K-V_T)^+ - \mathrm{e}^{\gamma T}V_0\mathsf{E}\left(\mathrm{e}^{\tilde{X}_T} - \tilde{K}\right)^+\mathbf{1}_{\{\tilde{M}_T<\log\tilde{K}\}}. \tag{4.8}$$

In some cases the expected loss can be negative, i.e., the expected return on the bond is higher than its face value K. Whether it is the case depends on the parameters of the process X_t and the constant γ.

From (4.7) we see that the pay–off of the bond in the model is equivalent to the pay-off of the portfolio consisting of risk–free K zero–coupon bonds, a short European put option with strike K, and a $\mathrm{e}^{\gamma T}V_0$ long European down-and-in call option on the exponential process X_t with strike $\tilde{K}$. Therefore the price of the defaultable bond at time 0 is

$$D(0,T) = KB(0,T) - PUT(0,T,K) + \mathrm{e}^{\gamma T}V_0 DIC(0,T,\tilde{K}), \tag{4.9}$$

where PUT is the price of the European put option and DIC is the price of the European down-and-in call option with time–varying barrier. Barrier option pricing in a Lévy processes framework is very hard and only partial results are known. A list of references can be found in Schoutens (2006).

In the original paper of Black and Cox (1976), the firm value process is assumed to follow a geometric Brownian motion, i.e.,

$$X_t = \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t.$$

Then

$$\tilde{X}_t = \left(\mu - \gamma - \frac{\sigma^2}{2}\right) t + \sigma W_t = \nu t + \sigma W_t,$$

where

$$\nu = \mu - \gamma - \frac{\sigma^2}{2}.$$

Using the reflection principle and Girsanov's theorem it is not hard to prove that the joint probability distribution of the Brownian motion $\tilde{X}_t$ and its running minimum $\tilde{M}_t$ satisfy

$$\mathsf{P}[\tilde{X}_t \geq x, \tilde{M}_t \geq y] = \Phi\left(\frac{-x+\nu t}{\sigma\sqrt{t}}\right) - \exp\left(\frac{2\nu y}{\sigma^2}\right)\Phi\left(\frac{2y-x+\nu t}{\sigma\sqrt{t}}\right). \tag{4.10}$$

The proof can be found in Musiela and Rutkowski (2005, Appendix B.4). Since

$$\mathsf{P}[\tilde{X}_t \geq x, \tilde{M}_t \geq y] = 1 - P[\tilde{X}_t < x] - P[\tilde{M}_t < y] + \mathsf{P}[\tilde{X}_t < x, \tilde{M}_t < y],$$

we can obtain the joint density function of $\tilde{X}_T$ and $\tilde{M}_T$ from (4.10) by differentiation. If we differentiate with respect to x we get

$$-\frac{1}{\sigma\sqrt{T}}\varphi\left(\frac{-x+\nu T}{\sigma\sqrt{T}}\right) + \frac{1}{\sigma\sqrt{T}}\exp\left(\frac{2\nu y}{\sigma^2}\right)\varphi\left(\frac{2y-x+\nu T}{\sigma\sqrt{T}}\right),$$

where φ is the density function of the standard normal distribution. Then by differentiating with respect to y we get the joint density function

$$f_{\tilde{X}_T,\tilde{M}_T}(x,y) = \frac{-2(2y-x)}{\sigma^3\sqrt{T^3}}\exp\left(\frac{2\nu y}{\sigma^2}\right)\varphi\left(\frac{2y-x+\nu T}{\sigma\sqrt{T}}\right).$$

The default probability (4.6) takes the form

$$\begin{aligned} DP &= \mathsf{P}[\tilde{M}_T < \log \tilde{K}] \\ &= 1 - \mathsf{P}[\tilde{M}_T \geq \log \tilde{K}] \\ &= 1 - \mathsf{P}[\tilde{M}_T \geq \log \tilde{K}, \tilde{X}_T \geq \log \tilde{K}] \\ &= 1 - \Phi\left(\frac{-\log \tilde{K} + \nu T}{\sigma\sqrt{T}}\right) + \tilde{K}^{\frac{2\nu}{\sigma^2}} \Phi\left(\frac{\log \tilde{K} + \nu T}{\sigma\sqrt{T}}\right) \\ &= \Phi\left(\frac{\log \tilde{K} - \nu T}{\sigma\sqrt{T}}\right) + \tilde{K}^{\frac{2\nu}{\sigma^2}} \Phi\left(\frac{\log \tilde{K} + \nu T}{\sigma\sqrt{T}}\right). \end{aligned}$$

Now let us compute the expected loss in the first–passage model. The first term in (4.8) is the same as the expected loss in Merton's model. For the valuation of the second term in (4.8) note that expression

$$\left(\mathrm{e}^{\tilde{X}_T} - \tilde{K}\right)^+ \mathbf{1}_{\{\tilde{M}_T < \log \tilde{K}\}}$$

is non–zero on the set

$$A = \left\{\mathrm{e}^{\tilde{X}_T} > \tilde{K}, \tilde{M}_T < \log \tilde{K}\right\} = \left\{\tilde{X}_T > \log \tilde{K}, \tilde{M}_T < \log \tilde{K}\right\}.$$

Therefore we can compute

$$\begin{aligned} &\mathsf{E}\left(\mathrm{e}^{\tilde{X}_T} - \tilde{K}\right)^+ \mathbf{1}_{\{\tilde{M}_T < \log \tilde{K}\}} \\ &\qquad = \int_A \left(\mathrm{e}^x - \tilde{K}\right) f_{\tilde{X}_T, \tilde{M}_T}(x, y)\, \mathrm{d}x\, \mathrm{d}y \\ &\qquad = \int_A \mathrm{e}^x f_{\tilde{X}_T, \tilde{M}_T}(x, y)\, \mathrm{d}x\, \mathrm{d}y - \tilde{K} \int_A f_{\tilde{X}_T, \tilde{M}_T}(x, y)\, \mathrm{d}x\, \mathrm{d}y. \end{aligned} \tag{4.11}$$

The first term in (4.11) is

$$\begin{aligned} &\int_A \mathrm{e}^x f_{\tilde{X}_T, \tilde{M}_T}(x, y)\, \mathrm{d}x\, \mathrm{d}y \\ &\qquad = \exp\left(\frac{2\nu \log \tilde{K}}{\sigma^2}\right) \int_{\log \tilde{K}}^{\infty} \mathrm{e}^x \frac{1}{\sqrt{2\pi\sigma^2 T}} \exp\left(-\frac{(-x + 2\log \tilde{K} + \nu T)^2}{2\sigma^2 T}\right) \mathrm{d}x \\ &\qquad = \exp\left(\frac{4\log \tilde{K}(\nu + \sigma^2) + \sigma^2 T(\sigma^2 + 2\nu)}{2\sigma^2}\right) \\ &\qquad\quad \times \int_{\log \tilde{K}}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2 T}} \exp\left(-\frac{(-x + 2\log \tilde{K} + \nu T + \sigma^2 T)^2}{2\sigma^2 T}\right) \mathrm{d}x \\ &\qquad = \tilde{K}^{\frac{2\nu + \sigma^2}{\sigma^2}} \mathrm{e}^{\frac{T}{2}(2\nu + \sigma^2)} \Phi\left(\frac{\log \tilde{K} + \nu T}{\sigma\sqrt{T}} + \sigma\sqrt{T}\right). \end{aligned}$$

The second term in (4.11) is equal to

$$\begin{aligned}
&\tilde{K}\mathsf{P}[\tilde{X}_T > \log\tilde{K}, \tilde{M}_T < \log\tilde{K}] \\
&\quad = \tilde{K}\left(\mathsf{P}[\tilde{X}_T > \log\tilde{K}] - \mathsf{P}[\tilde{X}_T > \log\tilde{K}, \tilde{M}_T \geq \log\tilde{K}]\right) \\
&\quad = \tilde{K}\left(1 - \Phi\left(\frac{\log\tilde{K} - \nu T}{\sigma\sqrt{T}}\right) - \Phi\left(\frac{-\log\tilde{K} + \nu T}{\sigma\sqrt{T}}\right) + \tilde{K}^{\frac{2\nu}{\sigma^2}}\Phi\left(\frac{\log\tilde{K} + \nu T}{\sigma\sqrt{T}}\right)\right) \\
&\quad = \tilde{K}^{\frac{2\nu+\sigma^2}{\sigma^2}}\Phi\left(\frac{\log\tilde{K} + \nu T}{\sigma\sqrt{T}}\right).
\end{aligned}$$

Putting it all together we get the expected loss

$$\begin{aligned}
\mathsf{E}\,L = {} & K\Phi(d_1) - V_0 \mathrm{e}^{\mu T}\Phi(d_1 - \sigma\sqrt{T}) \\
& - K\exp\left(\frac{2\nu}{\sigma^2}\left(\log\frac{K}{V_0} - \gamma T\right)\right)\left(\mathrm{e}^{\frac{T}{2}(2\nu+\sigma^2)}\Phi(d_2 + \sigma\sqrt{T}) - \Phi(d_2)\right),
\end{aligned}$$

where

$$\begin{aligned}
d_1 &= \frac{\log\frac{K}{V_0} - (\gamma + \nu)T}{\sigma\sqrt{T}}, \\
d_2 &= \frac{\log\frac{K}{V_0} - (\gamma - \nu)T}{\sigma\sqrt{T}}, \\
\nu &= \mu - \gamma - \frac{\sigma^2}{2}.
\end{aligned}$$

If we want to price a defaultable zero–coupon bond in the case of a geometric Brownian motion we can use (4.9), where for the pricing formulas of the European put option and especially the down-and-in call option we refer to Hull (2005) or Musiela and Rutkowski (2005).

A Step Further

The first–passage model is an extension of Merton's model and can be further extended by an excursion approach. Some authors argue that if the firm's value falls below the barrier D_t the creditor does not take control over the firm's assets immediately but he lets the firm reorganize and operate for a while. If the firm value does not rise, the creditor still takes control over the firm's assets. Therefore a default in this approach occurs after the firm value process spends some given time below the barrier. For the excursion approach overview see Giesecke (2004, Section 2.3). For more structural models and pricing of derivatives we refer to Bielecki and Rutkowski (2002).

4.4 Reduced–form Models

In reduced–form modeling the idea of a default is different than in structural models where the default depends on the firm's value. The default time is modeled here as some random variable τ using the theory explained in Chapter 2. The most common approaches of modeling the default time τ are through a Poisson process, an inhomogeneous Poisson process, or a Cox process (described in Section 4.4.1). The other possibility is an extension of default/non–default situation into credit ratings, which maps a credit worthiness of the debtor to some finite scale $1, \ldots, K$. This extension is described in Section 4.4.2.

Another difference in comparison with structural models is the fact that the recovery rates do not follow from default modeling implicitly and they need to be modeled separately. This feature gives us more freedom, which might be advantage same as disadvantage.

Since all models described in this chapter are mostly straightforward application of theory from Chapter 2 we strongly advised to read Chapter 2 before reading this chapter.

4.4.1 Default Probability

Let λ_t be the hazard rate of the default time τ and $N_t = \mathbf{1}_{\{\tau \leq t\}}$ the associated counting process. The hazard rate of the default time τ in the credit risk is often called a *default intensity*.

Poisson process In this model a hazard rate λ is constant and the default time τ has an exponential distribution with parameter λ. The exponential distribution is memory–less, which is not a very realistic assumption, but constant intensities are easy to estimate and to compute with. The default probability in the model is

$$DP_t = 1 - \mathsf{P}[N_t = 0] = 1 - \mathrm{e}^{-\lambda t}.$$

Time inhomogeneous Poisson process If one wants to fix the unrealistic assumption of constant default intensities he can assume a default intensity λ_t as a function of time. The time dependency can be estimated from historical data using some econometrics model (regression analysis) or can be given exogenously. The default probability is then

$$DP_t = 1 - \mathsf{P}[N_t = 0] = 1 - \exp\left(-\int_0^t \lambda_s \,\mathrm{d}s\right).$$

Cox process Finally, we incorporate the future uncertainty and assume stochastic default intensities. In the Cox process setting it is assumed that there exists some risk driver factor X_t which is often modeled via *affine models* (Section 2.1.6). The default intensity λ_t is then a deterministic function of X_t, i.e., $\lambda_t = \lambda(X_t)$. Thus, conditioned on the realization of X_t we get an inhomogeneous Poisson process. The default probability in the Cox model is then

$$DP_t = 1 - \mathsf{P}[N_t = 0] = 1 - \mathsf{E}\left[\exp\left(-\int_0^t \lambda(X_s) \,\mathrm{d}s\right)\right].$$

Often a choice for the function $\lambda(x)$ is $\lambda(x) = x$. Furthermore, if we assume that the risk factor X_t follows some affine process, we can get a closed–form expression of the form

$$1 - \exp\big(A(t) - B(t)X_0\big),$$

for some $A(t)$ and $B(t)$. For example, if we assume the model of Vašíček (1977) as in Section 2.1.6, we have

$$DP_t = 1 - \exp\big(A(t) - B(t)X_0\big),$$

where

$$B(t) = \frac{1}{\kappa}\left(1 - \mathrm{e}^{-\kappa t}\right),$$
$$A(t) = \frac{(B(t) - t)(\kappa^2\theta - \frac{\sigma^2}{2})}{\kappa^2} - \frac{\sigma^2 B(t)^2}{4\kappa}.$$

Another possibility is to use time subordination which we discussed in Section 3.1.2.

4.4.2 Credit Rating Migration

Sometimes it is useful to describe the credit worthiness of the obliger in more detail than just a default or a non–default. A way to do this is to introduce *credit ratings*. A credit rating is a grade on a finite scale $\{1, \dots, K\}$ expressing the credit worthiness of the obliger. The rating 1 is the best and the rating K is a default. The default rating K is absorbing, i.e., once the default state is reached it lasts until the end of the assumed horizon. As a special case when $K = 2$ we get a model where just a default and a non–default are assumed (as in the previous scope). In some credit rating systems there might be more credit ratings for a default.[21] We will speak about default only in case of rating K, if not stated otherwise. The *rating agency* is a company which measures the credit worthiness of the obliger and then assigns a credit rating from its rating systems. Probably the most well–known agencies are Standard&Poors (S&P) and Moody's. These agencies assign credit rating only to big international companies or to states. Hence many banks developed their own rating systems which they use for internal purposes.

If the bank uses its own internal rating system it knows the present credit worthiness of the obliger, but it also needs to model the future development of the obliger's credit worthiness, i.e., how the obliger will migrate between different ratings. The credit rating migration is often modeled using Markov chains as was introduced in Jarrow et al. (1997). The credit rating process is assumed to be a discrete–time or a continuous–time Markov chain. It is quite expensive for a bank to update the credit rating for all their obligers continuously and hence they update it semi–annually in most cases. Therefore, it makes sense to model the credit rating process as a discrete Markov chain. From the other point of view, the credit worthiness of the obliger changes continuously and assigning ratings on a semi–annual base is just a discrete observation of a continuous process and hence one should use a continuous–time Markov chain. The continuous approach is also more tractable and uses more information about transitions than the discrete–time Markov chain.

The continuous–time homogeneous Markov chain model has a few not very realistic features that real data do not approve.

1. Constant rating intensities — Real data shows that intensities change over time.
2. The exponential distribution of sojourn times — The exponential distribution is memoryless, but data shows some rating "momentum", i.e., the time to the next rating change depends on the time already spend in a current rating.
3. The Markov property — Transition probabilities should depend only on current rating, but empirically there is evidence that after a rating downgrade, there is a higher probability of another downgrade than if a current rating was reached by an upgrade.

For more studies see Carty and Fons (1993), Kavvathas (2001) and Lando and Skødeberg (2002).

The last mentioned problem can be solved by extending the state space from K possible ratings $\{1, \dots, K\}$ to $2K - 2$ in the following way

$$1, 2, 2-, 3, 3-, \dots, (K-1)-, K,$$

where rating k (resp. $k-$) means that rating k was reached by an upgrade (resp. downgrade). Note that state 1 can not be reached by a downgrade and state K can be reached only by a downgrade, hence $2K - 2$ states. After this extension we can renumber ratings and work with the state space $\{1, \dots, 2K - 2\}$. A disadvantage of the method is that we need more data for

[21] For example, 3 rating grades are default and they differ from each other in recovery rates. The first one assume a constant recovery of 66 %, the second one of 33 % and the last one has 0 recovery.

estimation of transition probabilities. Note that the number of transition probabilities, which we need to estimate, rises with the second power of the number of states. In practice there is often not enough data and some ratings are joining together for lower the data demand for estimation.

The other two problems of a time–homogeneous Markov chains can be solved by introducing a time–inhomogeneous continuous Markov chain, time homogeneous semi–Markov chain, a time–inhomogeneous semi–Markov chain, or even transition intensities that are model by stochastic processes.

In general, we are interested in modeling transition probabilities

$$\mathbf{P}(s,t) = \Big(p_{ij}(s,t)\Big)_{i,j=1}^{K},$$

where $p_{ij}(s,t)$ is the probability that the debtor will be in rating j at time t conditioned that he is in rating i at time s. We assume that there exists a $K \times K$-dimensional process $\mathbf{Q}(t)$ such that

$$\mathbf{P}(s,t) = \exp\left(\int_s^t \mathbf{Q}(u)\,\mathrm{d}u\right),$$

if $\mathbf{Q}(t)$ is a deterministic matrix function of time and

$$\mathbf{P}(s,t) = \mathsf{E}\left[\exp\left(\int_s^t \mathbf{Q}(u)\,\mathrm{d}u\right)\right],$$

if $\mathbf{Q}(t)$ is a stochastic process. In both cases the integral is assumed componentwise and we assume that these integrals exist and for every t, $\mathbf{Q}(t)$ is a valid generator.

The last column of the matrix $\mathbf{P}(s,t)$ is a vector of default probabilities for different ratings. Let us denote it by

$$\boldsymbol{DP}(s,t) = \big(DP_i(s,t)\big)_{i=1}^{K-1},$$

where

$$DP_i(s,t) = p_{iK}(s,t),$$

is the probability that the debtor will default up to time t conditioned he is in rating i at time s. The probability matrix $\mathbf{P}(t)$ can be written in the form

$$\mathbf{P}(s,t) = \begin{pmatrix} \tilde{\mathbf{P}}(s,t) & \boldsymbol{DP}(s,t) \\ \boldsymbol{0} & 1 \end{pmatrix}, \tag{4.12}$$

where the matrix $\tilde{\mathbf{P}}(s,t)$ is a $(K-1)\times(K-1)$ matrix and $\boldsymbol{0}$ and $\boldsymbol{DP}(s,t)$ are $(K-1)$ dimensional vectors. Since $\mathbf{P}(s,t)$ is a stochastic matrix it holds for every $0 \le t_1 \le t_2 \le \ldots \le t_n$ that

$$\mathbf{P}(t_1,t_n) = \mathbf{P}(t_1,t_2)\cdot\mathbf{P}(t_2,t_3)\cdot\ldots\cdot\mathbf{P}(t_{n-1},t_n). \tag{4.13}$$

Using (4.12) and (4.13) it follows the intuitive relation

$$\boldsymbol{DP}(t_1,t_n) = \boldsymbol{DP}(t_1,t_2) + \tilde{\mathbf{P}}(t_1,t_2)\boldsymbol{DP}(t_2,t_3) + \ldots + \tilde{\mathbf{P}}(t_1,t_{n-1})\boldsymbol{DP}(t_{n-1},t_n),$$

which is useful if we are interested in some particular discrete time evolution of default probabilities.

If we use a time–homogeneous continuous–time Markov chain (CTMC) we can compute transition probabilities and especially default probabilities by very straightforward application of the theory from Section 2.2.2. Transition probability matrix is

$$\mathbf{P}(s,t) = \exp\big((t-s)\mathbf{Q}\big) = \sum_{n=0}^{\infty} \frac{(t-s)^n \mathbf{Q}^n}{n!}, \qquad t \geq s \geq 0.$$

Time–homogeneous continuous–time Markov chain are the most commonly used in practice. The only thing which we need to be able to use CTMC model is generator matrix $\mathbf{Q}$. This matrix is usually estimated from historical data using ML method (Section 2.2.2). However, as we mentioned above banks usually do not have continuous–time observation of credit ratings of debtors and continuous–time observation is one of the main assumptions of ML estimation. Hence we rather recommend to use one of the methods for partially observed data from Section 2.3.

A direct consequence of using time–homogeneous model is a dynamics which does not change over time, which is denied by empirical studies. If we want to allow dynamics changes over time we could use a more complicated structure of the inhomogeneous continuous–time Markov chain, where we need to specify the structure of inhomogeneity and justify it by some arguments. That is not easy at all. Bluhm and Overbeck (2007) suggest to use an inhomogeneous continuous–time Markov chain with the generator

$$\mathbf{Q}_t = \mathbf{D}(t)\mathbf{Q}, \qquad t \geq 0,$$

where $\mathbf{Q}$ is a constant valid generator and $\mathbf{D}(t)$ is a diagonal matrix with diagonal elements

$$d_{ii} = \frac{\big(1 - \mathrm{e}^{-\alpha_i t}\big) t^{\beta_i - 1}}{1 - \mathrm{e}^{-\alpha_i}}, \qquad i = 1, \ldots, K,$$

and $\alpha_i, \beta_i > 0$ are non–negative parameters of the model. Note that the generator $\mathbf{Q_t}$ came up from constant $\mathbf{Q}$ by multiplying the i-th row by d_{ii}. Using this inhomogeneity structure they reached a significantly better fit to default probabilities on data from S&P. Bluhm and Overbeck (2007) also provide some discussion, where they pointed out that the choice of the structure is not completely random.

The main result of this paper is the usage of time–changed continuous–time Markov chains described in Chapter 3, which is simpler than the structure of Bluhm and Overbeck (2007) and have a possible interpretation that system runs homogeneously with respect to some business time. Mapping of this business time can be deterministic or random. Modeling of the business time give us easy possibility how to influence the system dynamics through only one parameter. When business time is faster than a calendar time, economy runs faster and debtors default faster (more defaults in given time interval) and contrary when business time is slower than a a calendar time, economy runs faster and debtors default slower (less defaults in given time interval). It make sense that rules and dynamics of economy does not change, but only runs faster or slower. In Chapter 5 we show application of our methods on real data with significantly better results than with traditional methods.

Let us assume that we are given by time evolution T_t then transition probability matrix is

$$\mathbf{P}(s,t) = \exp\big((T_t - T_s)\mathbf{Q}\big) = \sum_{n=0}^{\infty} \frac{(T_t - T_s)^n \mathbf{Q}^n}{n!}, \qquad t \geq s \geq 0.$$

If time evolution T_t is random transition probability matrix is then[22]

$$\mathbf{P}(s,t) = \mathsf{E}\left[\exp\left((T_t - T_s)\mathbf{Q}\right)\right] = \sum_{n=0}^{\infty} \frac{\mathsf{E}\left[(T_t - T_s)^n\right]\mathbf{Q}^n}{n!}, \qquad t \geq s \geq 0,$$

Recall that the deterministic time evolution is a special case of random time evolution. In Chapter 3 we described how the time–changed system behaves and how to estimate parameters in three possibly approaches:

1. Time evolution is given exogenously.
2. Time evolution is deterministic but unknown.
3. Time evolution is random variable from distribution with some unknown parameters.

In each of these approaches we derived ML estimators and also extended some methods in case of partially observed data.

Big advantage of the time change is that we can model it separately of the loan portfolio and build a model for predicting time evolution which depends on a macroeconomic variables. In Chapter 5 we estimated a time evolution from data and than we computed correlation coefficient of estimated time evolution and FFTR (federal fund target rate), USA GDP growth and S&P500 index growth. Despite the fact that data about defaults are worldwide and FFTR, GDP growth and S&P500 index are mainly indicators of USA economy, the correlation is around 60%. That suggests that modeling and predicting the time evolution by some econometric model depending on macroeconomic variables can give us very good estimates of time evolution. These good estimates of time evolution can be used for default probability modeling.

4.4.3 Recovery Rates

The essential part of the risk management modeling are recovery rates RR. Unlike in structural models recovery rates do not follow implicitly from reduced–form models. The big problem of the recovery modeling is a lag of data for estimation. Only some banks have sufficient historical data about recoveries. Therefore some simplifying assumptions are necessary to enable any modeling.

Recall that the loss process of the defaultable zero–coupon bond with face value 1 has dynamics

$$L_t = (1 - RR_\tau)\mathbf{1}_{\{\tau \leq t\}}.$$

Then we can compute the expected loss in general as

$$\mathsf{E}\, L_t = \mathsf{E}\left[(1 - RR_\tau)\mathbf{1}_{\{\tau \leq t\}}\right] = \mathsf{E}\left[\mathsf{E}\left[(1 - RR_\tau)\mathbf{1}_{\{\tau \leq t\}}|\tau\right]\right] = \mathsf{E}\left[\mathbf{1}_{\{\tau \leq t\}}\,\mathsf{E}\left[(1 - RR_\tau)|\tau\right]\right].$$

The recovery rate RR_t can be a deterministic function of time and then RR_τ is a random variable because it is a value of the function in random time. In general, the recovery RR_t can be a stochastic process with values in $[0,1]$.[23]

If recovery rates are independent of the random default time τ, i.e., the random variable $RR_t = RR$ has the same distribution for every t, then

$$\mathsf{E}\, L_t = DP_t\,\mathsf{E}\,[1 - RR].$$

[22] We assume we can interchange integration and summation.

[23] Rarely we can allow recovery rates higher than 1.

In literature for pricing credit derivatives there are three main approaches to recoveries. We will mention each approach and compute the expected loss of the bond in this model when we assume a Cox process as the driver of the default time τ.

Recovery of Face Value (RFV) — The recovery is assumed to be the exogenously given fraction c of the face value of the bond. Then recovery rates are constant and independent of τ

$$RR_\tau = c,$$

and the expected loss is then

$$\mathsf{E}\, L_t = \mathsf{E}\, \left[\mathbf{1}_{\{\tau\le t\}}\, \mathsf{E}\, [(1-c)|\tau]\right] = (1-c)\mathsf{E}\, \mathbf{1}_{\{\tau\le t\}} = (1-c)DP_t.$$

Recovery of Treasury (RT) — The recovery is assumed to be the exogenously given fraction c of an equivalent risk free zero–coupon bond. Then we need to specify dynamics of the risk free zero–coupon bonds. The recovery rate is a random variable

$$RR_\tau = cB(\tau, T),$$

and the expected loss is then

$$\mathsf{E}\, L_t = \mathsf{E}\, \left[\mathbf{1}_{\{\tau\le t\}}\, \mathsf{E}\, [\big(1-cB(\tau,T)\big)|\tau]\right] = DP_t - c\,\mathsf{E}\, [\mathbf{1}_{\{\tau\le t\}}B(\tau,T)].$$

In the case when τ is exponentially distributed with parameter λ and the risk free interest rate r is constant, we have

$$\begin{aligned}
RR_\tau &= c\mathrm{e}^{-r(T-\tau)}\\
\mathsf{E}\, L_t &= DP_t - c\mathrm{e}^{-rT}\mathsf{E}\, \left[\mathbf{1}_{\{\tau\le t\}}\mathrm{e}^{r\tau}\right]\\
&= 1-\mathrm{e}^{-\lambda t} - c\mathrm{e}^{-rT}\int_0^\infty \mathbf{1}_{\{x\le t\}}\,\mathrm{e}^{rx}\,\lambda\mathrm{e}^{-\lambda x}\,\mathrm{d}x\\
&= 1-\mathrm{e}^{-\lambda t} - \frac{\lambda c}{\lambda-r}\mathrm{e}^{-rT}\int_0^t (\lambda-r)\mathrm{e}^{-(\lambda-r)x}\,\mathrm{d}x\\
&= 1-\mathrm{e}^{-\lambda t} - \frac{\lambda c}{\lambda-r}\mathrm{e}^{-rT}\left(1-\mathrm{e}^{-(\lambda-r)t}\right).
\end{aligned}$$

Recovery of Market Value (RMV) — The recovery is assumed to be an exogenously given fraction of the pre–default market value of the bond, i.e.,

$$RR_\tau = cD(\tau_-, T).$$

The expected loss is then

$$\mathsf{E}\, L_t = DP_t - c\,\mathsf{E}\, [\mathbf{1}_{\{\tau\le t\}}D(\tau_-,T)].$$

For further computation we need to know the dynamics of the defaultable zero–coupon bonds and how to price them. Pricing of defaultable coupon bonds is a very wide topic and beyond the scope of this work. We refer to Duffie and Singleton (1999) and Collin-Dufresne et al. (2004) for that.

In all three cases we mentioned the constant c. This constant is estimated from historical data. We can go further and assume that the fraction c is not constant anymore but is a random variable C with values on $[0,1]$.

For purpose of modeling the random fraction C, we could use the uniform distribution on $[0,1]$, but empirical studies show the assumption of the uniform distribution to be very unrealistic. In Schuermann (2004) there are some empirical distributions of recoveries of American bonds and loans. These empirical distributions are highly right–skewed.

A much better distribution than the uniform distribution is the very flexible Beta distribution with density function

$$f(x) = \frac{1}{\mathbf{B}(a,b)} x^{a-1}(1-x)^{b-1},$$

where $a, b > 0$ and $\mathbf{B}$ is the Beta function defined as

$$\mathbf{B}(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x.$$

The parameters a and b can be estimated from data using the moment method or the maximum likelihood estimation. Note that the uniform distribution is special kind of the Beta distribution with parameters $a = b = 1$.

The Beta distribution has mean value

$$\frac{a}{a+b}.$$

Therefore if we assume the RFV approach we get the expected loss

$$\mathsf{E}\, L_t = \left(1 - \frac{a}{a+b}\right) DP_t.$$

Dependent Recoveries

Some empirical studies indicate that recovery rates and default intensities are not independent, but they tend to be negatively correlated. They show that in the recession of the economy, default intensities tend to be higher and recoveries lower.

One of the usual ways of modeling a dependent recovery and a default intensity λ_t is to introduce an additional possibly multi–dimensional process X_t that can be interpreted as a state of the economy. The random fraction C and the default intensity λ_t both depend on X_t, but conditioned on the realization X_t they are independent of each other.

In our Beta distribution setting, it can be, for example, such that parameters a and b are deterministic functions of X_t. If we assume the RFV approach we have the expected loss

$$\begin{aligned}
\mathsf{E}\, L_t &= \mathsf{E}\left[\big(1 - C(X_t)\big)\mathbf{1}_{\{\tau \le t\}}\right] \\
&= \mathsf{E}\left[\mathsf{E}\left[\big(1 - C(X_t)\big)\mathbf{1}_{\{\tau \le t\}} | X_t\right]\right] \\
&= \mathsf{E}\left[\big(1 - \mathsf{E}\left[C(X_t)|X_t\right]\big)\,\mathsf{E}\left[\mathbf{1}_{\{\tau \le t\}}|X_t\right]\right].
\end{aligned}$$

In RT and RMV models things become more complicated since for these approaches we need a dynamics of interest rates for a zero coupon bond modeling, and interest rates can be correlated with default intensities and recoveries. The interest rate is also often taken as a dimension of the process X_t. Sometimes even the process X_t is chosen as $X_t = r_t$.

An interesting idea is also applying our random time evolution extension, where we set $X_t = T_t$. The intuition behind this setting is very similar to the credit ratings modeling (faster or slower business time) and we get quite clear connection between the credit migration modeling and the recovery rates modeling. For finding any working model we need data about

credit ratings migration and related recovery rates, which we unfortunately do not have. Hence we are not able to propose any possible working model.

We should also mention the dependency of recovery rates on the character of the debtors business. Imagine that we have two debtors. The first one is a real estate agent with only one debt and a big equity and the second one is a financial derivative trader with a minimal equity. One could hardly expect that recovery rates of both of them have the same probability distribution. Therefore we can divide debtors to some groups and estimate the distribution of recovery rates within these groups. By dividing into groups we are limited by accessible historical data.

4.5 Incomplete Information Models

In Section 4.3 we have described structural models in a general Lévy processes setting, but in most cases just a Brownian motion as the risk driver is used. As a corollary of the continuity of a Brownian motion paths there is a vanishing credit spread for short term debts, because it is very unlikely for a continuous process to steeply decrease in a short period. Furthermore, due to the continuity, the default time is a predictable stopping time.

A pioneering work in incomplete information models is Duffie and Lando (2001), where they argue that the firm's value process is not observable directly and continuously to public, but it is observable to the management of the firm. The information available to company managers forms the filtration $\mathcal{F}_t$ which is different from the filtration $\mathcal{G}_t$ which is formed by the information available to the market.

Duffie and Lando (2001) assumed that the information from filtration $\mathcal{F}_t$ is revealed to public ("transferred" to filtration $\mathcal{G}_t$) in some discrete times (for example earnings announcement) and even then it is revealed with some additional noise. A default time τ is then a predictable stopping time with respect to the filtration $\mathcal{F}_t$, but is a totally inaccessible stopping time[24] with respect to the filtration $\mathcal{G}_t$. This idea was motivated by accounting scandals in American companies Enron and WorldCom. Both of them had mistakes in their accounting which is by Duffie and Lando (2001) viewed as an additional noise.

Kusuoka (1999) assumes similar to Duffie and Lando (2001) that information is revealed to the market with an additional noise, but in a continuous way.

In Collin-Dufresne et al. (2003) the authors choose a different approach. They assume that the information about the company is revealed to the public with some delay, but continuously and completely.

Çetin et al. (2004) suggest an approach in which the information available to the management is not completely revealed to public, but just some fraction of it.

Giesecke (2006) described a first–passage model where the default barrier is random. That is also the incomplete information approach.

Guo et al. (2008) try to present a unified framework for incomplete information models using the delayed filtration. Especially, they reconcile the approach by Duffie and Lando (2001) and Collin-Dufresne et al. (2003).

Behind every credit risk model (it does not matter if it is structural, reduced–form or an incomplete information model), there is a counting process $N_t = \mathbf{1}_{\{\tau \leq t\}}$ for which according to the Doob–Mayer decomposition there exists a compensator A_t such that $N_t - A_t$ is a martingale. The compensator A_t is often called a *default drift* in this context. A reduced–form model assumes the existence of the default intensity and the default intensity is a major tool in

[24] A stopping time τ is totally inaccessible if and only if the probability that τ is a limit of any sequence of predictable stopping times τ_n is zero.

reduced–form models. An incomplete information model uses rather directly the compensator process A_t than the intensity itself. If the intensity exists the model is unified with a reduced–form model. So incomplete information models can be seen as a generalization of a reduced–form model based on structural model ideas. The default probability is then under some technical conditions equal to

$$DP_t = 1 - \mathsf{E}\left[\mathrm{e}^{-A_t}\right].$$

For further details about particular incomplete information model see the article referenced above.

4.6 Dependent Default Modeling

There is no need to investigate the dependency between particular debtors in the expected loss modeling, since the expected loss is just a sum of expected losses on particular loans. But if we want to compute the variance of the credit loss of the portfolio, dependencies between debtors comes into play as a major factor.

Let us assume two big companies from the same industry and location, and one of them has defaulted. Then one could expect that the default probability of the second company will raise, since both companies are influenced by very similar risk factors on the market.

Recall that the one–year credit loss of the portfolio is given by

$$L = \sum_{i=1}^{n} EAD \times LGD \times \mathbf{1}_{\{\tau_i \leq 1\}}.$$

For demonstrating the idea let us assume that $EAD = LGD = 1$, then

$$L = \sum_{i=1}^{n} \mathbf{1}_{\{\tau_i \leq 1\}}.$$

The expected one–year loss is then simply

$$\mathsf{E}\, L = \sum_{i=1}^{n} \mathsf{E}\, \mathbf{1}_{\{\tau_i \leq 1\}} = \sum_{i=1}^{n} DP_1^i,$$

where DP_1^i is the one–year default probability on the i-th loan. The different situation is if we want to compute the variance of the credit loss

$$\begin{aligned} \mathsf{Var}\,(L) &= \mathsf{Var}\,\left(\sum_{i=1}^{n} \mathbf{1}_{\{\tau_i \leq 1\}}\right) \\ &= \sum_{i=1}^{n} \mathsf{Var}\,\left(\mathbf{1}_{\{\tau_i \leq 1\}}\right) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathsf{Cov}\,\left(\mathbf{1}_{\{\tau_i \leq 1\}}, \mathbf{1}_{\{\tau_j \leq 1\}}\right), \end{aligned} \tag{4.14}$$

where

$$\begin{aligned} \mathsf{Cov}\,\left(\mathbf{1}_{\{\tau_i \leq 1\}}, \mathbf{1}_{\{\tau_j \leq 1\}}\right) &= \mathsf{E}\,\left[\mathbf{1}_{\{\tau_i \leq 1\}}, \mathbf{1}_{\{\tau_j \leq 1\}}\right] - \mathsf{E}\,\mathbf{1}_{\{\tau_i \leq 1\}} \mathsf{E}\,\mathbf{1}_{\{\tau_j \leq 1\}} \\ &= \mathsf{P}[\tau_i \leq 1, \tau_j \leq 1] - DP_i DP_j. \end{aligned}$$

Also for other statistics like VaR or CVaR we need to know how default events of particular debtors depend on each other. As a matter of fact we should go even further and model the

joint distribution of defaults and recoveries, but it is beyond the scope of the presented work. In the following sequel we assume independent recovery rates and default events unless stated otherwise.

Let us imagine the situation when the bank loan portfolio consists of n loans and a new $(n+1)$-st loan is considered. What will be a contribution of a new loan to the current portfolio loss variance? From (4.14) we see that the contribution will be

$$\mathsf{Var}\left(\mathbf{1}_{\{\tau_{n+1}\leq 1\}}\right) + 2\sum_{i=1}^{n} \mathsf{Cov}\left(\mathbf{1}_{\{\tau_i\leq 1\}}, \mathbf{1}_{\{\tau_{n+1}\leq 1\}}\right),$$

which can be even negative if the default event of the new debtor is negatively correlated with default events of current debtors. In that case it is very desirable to accept this loan, since it diversifies the risk.

4.6.1 Unit Loss and Homogeneous Portfolio

The larger the loan portfolio is, the harder and more time consuming the calculations are. It is also more complicated to estimate all necessary parameters (default probabilities, covariances, recovery rates,...) for every obliger. Hence a concept of the *homogeneous portfolio* is often assumed. It is an assumption that exposure, default probability, recovery rates and also dependencies are the same all over the portfolio.[25] Without loss of generality we can take exposure equal to 1. Then, the number of defaults divided by the number of loans in the portfolio is the fraction of the whole portfolio exposure, which will be lost.

4.6.2 Factor Models

As we mentioned above, two companies doing similar business are sensitive to similar risk drivers. Also one can imagine that the location of these two companies play a significant role (risk factor) for the default event. Finally, let there be an economy recession and people are not willing to spend much money. It will influence many different businesses in many different ways and we would like to know how.

The *factor model* is a dependency structure of a particular firm risk on some common risk factors such as a business factor, country factor, etc. Part of the firm's riskiness that can not be explained by common risk factors is called an *idiosyncratic risk* or *residual risk*.

Let $\boldsymbol{Z}_t$ be a d-dimensional risk factor common to all firms in the portfolio. Let us take an arbitrary firm i. Now we are interested in a way and how much riskiness of the firm i depends on the risk factor $\boldsymbol{Z}_t$.

Let us assume that we use a structural model. Recall that we assume that the firm value process of firm i has dynamics

$$V_t^{(i)} = V_0^{(i)} \exp\left(X_t^{(i)}\right),$$

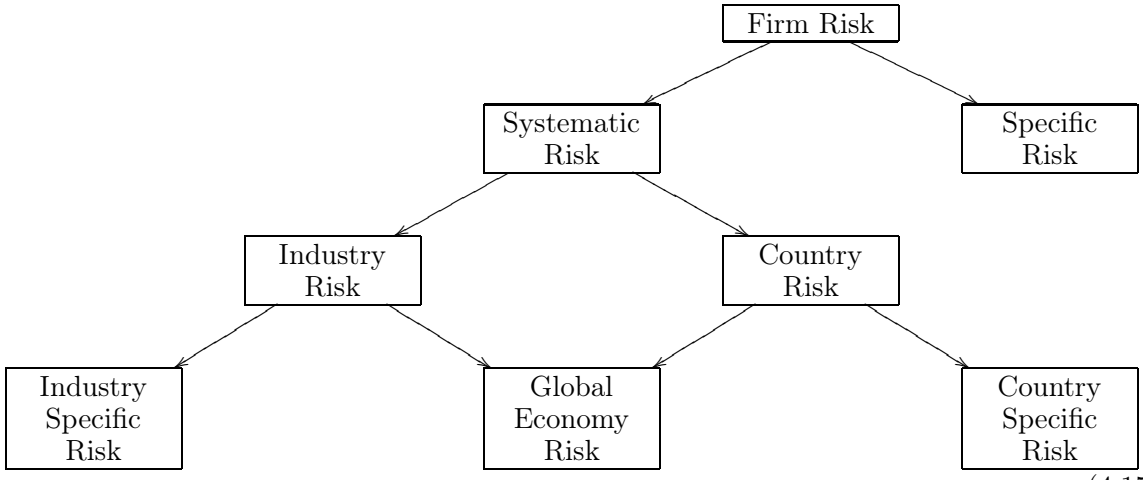
where $X_t^{(i)}$ is some Lévy process. In factor models we assume that the risk driver $X_t^{(i)}$ is a linear function of the common risk factor $\boldsymbol{Z}_t$ and some idiosyncratic process $\varepsilon_t^{(i)}$ which is specific to firm i, i.e.,

$$X_t^{(i)} = \boldsymbol{\beta}^{(i)} \cdot \boldsymbol{Z}_t + \varepsilon_t^{(i)},$$

[25] The homogeneity assumption is useful especially in the retail segment, where a lot of very similar loans exist and they behave in a similar pattern. The bigger loans and smaller number of loans (big corporate loans), the less accurate results models based on the homogeneous assumption will give.

where processes $\boldsymbol{Z}_t, \varepsilon_t^{(1)}, \ldots, \varepsilon_t^{(m)}$ are independent of each other. Since the process $\boldsymbol{Z}_t$ is common to all firms in the economy (or portfolio), default events are dependent if $\boldsymbol{\beta}^{(i)} \neq \boldsymbol{0}$. Conditionally on the realization of the common risk factor $\boldsymbol{Z}_t$ processes $X_t^{(1)}, \ldots, X_t^{(n)}$ are independent. This fact is often used in computations or simulations. The d-dimensional real parameter $\boldsymbol{\beta}^{(i)}$ is estimated from historical data using linear regression methods and is often called *loads*.

The common risk factor can have, for example, the structure as in diagram (4.15) where the three–level factor model of KMV[26] is described as presented in Bluhm et al. (2002, Figure 1.7).



(4.15)

We see that the riskiness of the firm is decomposed into three common factors (industry specific risk, global economy risk, and country specific risk) and a firm specific risk (idiosyncratic risk).

The special case of factor models is the *one–factor model* where $\boldsymbol{Z}_t$ is one–dimensional. This model is very often used in practice since it is very simple.

In reduced–form modeling a factor model can be used to decompose intensities $\lambda_t^{(i)}$. Then we have

$$\lambda_t^{(i)} = \boldsymbol{\beta}^{(i)} \cdot \boldsymbol{Z}_t + \varepsilon_t^{(i)}, \tag{4.16}$$

where $\varepsilon_t^{(i)}$ are independent of $\boldsymbol{Z}_t$ and each other. In a reduced form model λ_t needs to be non–negative. It can be done by truncating $\lambda_t^{(i)}$, which comes from model (4.16).

For further discussion about factor models, see Bluhm et al. (2002, Section 1.2.3).

4.6.3 Bernoulli Models

Let us fix the time horizon on one year and let us assume a unit exposure and zero recovery on every loan in the portfolio. Then the aggregate credit loss is

$$L = \sum_{i=1}^{n} \mathbf{1}_{\{\tau_i \leq 1\}}.$$

We will describe a deterministic Bernoulli model and its extension to the general Bernoulli mixture model.

[26] KMV is a part of Moody's and is one of the biggest provider of credit risk solutions in world.

"Deterministic" Bernoulli Model

In the Bernoulli model we think about a default event $Y_i = \mathbf{1}_{\{\tau_i \le 1\}}$ as about a Bernoulli random variable which equals 1 with probability $p_i = DP_i$ and 0 with probability $1 - p_i$, where DP_i is the one–year default probability of loan i. Furthermore, default events are assumed to be independent of each other. Then we get

$$\begin{aligned} \mathsf{E}\, L &= \sum_{i=1}^{n} \mathsf{E}\, Y_i = \sum_{i=1}^{n} p_i, \\ \mathsf{Var}\,(L) &= \sum_{i=1}^{n} \mathsf{Var}\,(Y_i) = \sum_{i=1}^{n} p_i(1-p_i). \end{aligned}$$

In the case of the homogeneous portfolio, where we assume the same default probability DP on every loan, the portfolio loss L is the binomial random variable with parameters n and $p = DP$. Thus,

$$\begin{aligned} \mathsf{E}\, L &= n\, p, \\ \mathsf{Var}\,(L) &= n\, p(1-p). \end{aligned}$$

General Bernoulli Mixture Model

Independent default events is not very realistic assumption, hence we extend the model to a general Bernoulli mixture model. Default events are still the Bernoulli random variables with parameter $P_i = DP_i$, but the parameter P_i is not deterministic anymore. Parameters $(P_1, \ldots, P_n)$ are random variables with the joint n-dimensional distribution $\boldsymbol{F}$ on $[0,1]^n$. Conditioned on the realization $(p_1, \ldots, p_n)$ of $(P_1, \ldots, P_n)$, the default event Y_i is a Bernoulli random variable with parameter p_i independent of $Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n$. Then it follows that

$$\mathsf{P}[Y_1 = y_1, \ldots, Y_n = y_n] = \int_{[0,1]^n} \prod_{i=1}^{n} p_i^{y_i} (1-p_i)^{1-y_i}\, \mathrm{d}\boldsymbol{F}(p_1, \ldots, p_n), \tag{4.17}$$

where $y_i \in \{0, 1\}$. The expected loss is then

$$\mathsf{E}\, L = \sum_{i=1}^{n} \mathsf{E}\,[Y_i] = \sum_{i=1}^{n} \mathsf{E}\,\left[\mathsf{E}\,[Y_i|P_i]\right] = \sum_{i=1}^{n} \mathsf{E}\,[P_i].$$

Before we will compute the variance of the credit loss note that

$$\begin{aligned} \mathsf{Var}\,(Y_i) &= \mathsf{Var}\,\left(\mathsf{E}\,[Y_i|P_i]\right) + \mathsf{E}\,\left[\mathsf{Var}\,(Y_i|P_i)\right] \\ &= \mathsf{Var}\,(P_i) + \mathsf{E}\,\left[P_i(1-P_i)\right] \\ &= \mathsf{E}\,[P_i^2] - \left(\mathsf{E}\,[P_i]\right)^2 + \mathsf{E}\,[P_i] - \mathsf{E}\,[P_i]^2. \\ &= \mathsf{E}\,[P_i]\left(1 - \mathsf{E}\,[P_i]\right), \end{aligned} \tag{4.18}$$

and

$$\begin{aligned} \mathsf{Cov}\,(Y_i, Y_j) &= \mathsf{E}\,[Y_i Y_j] - \mathsf{E}\,[Y_i]\mathsf{E}\,[Y_j] \\ &= \mathsf{E}\left[\mathsf{E}\,[Y_i Y_j|P_i, P_j]\right] - \mathsf{E}\,[P_i]\mathsf{E}\,[P_j] \\ &= \mathsf{E}\,[P_i P_j] - \mathsf{E}\,[P_i]\mathsf{E}\,[P_j] \\ &= \mathsf{Cov}\,(P_i, P_j). \end{aligned}$$

We see that the whole correlation structure of default events is equipped by the correlation structure of random parameters $P_i, i = 1, \ldots, n$. The variance of the aggregate credit loss then follows from (4.14)

$$\mathsf{Var}\,(L) = \sum_{i=1}^{n} \mathsf{E}\,[P_i]\,(1 - \mathsf{E}\,[P_i]) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathsf{Cov}\,(P_i, P_j).$$

Homogeneous Portfolios In the case of a homogeneous credit portfolio, the default probability on every loan is the same. Therefore, we have just one random parameter P with probability distribution F. We can rewrite (4.17) into the form

$$\mathsf{P}[Y_1 = y_1, \ldots, Y_n = y_n] = \int_0^1 p^k (1-p)^{n-k}\,\mathrm{d}F(p),$$

where $k = \sum_{i=1}^{n} y_i$, $y_i \in \{0, 1\}$, is the number of defaults in the portfolio. The probability that exactly k defaults occur is then

$$\mathsf{P}[L = k] = \binom{n}{k} \int_0^1 p^k (1-p)^{n-k}\,\mathrm{d}F(p).$$

Let us denote the uniform default probability by $\bar{p}$. Then

$$\bar{p} = \mathsf{P}[Y_i = 1] = \mathsf{E}\,[Y_i] = \mathsf{E}\,[P] = \int_0^1 p\,\mathrm{d}F(p),$$

further using (4.18) we get

$$\begin{aligned}
\mathsf{Corr}\,(Y_i, Y_j) &= \frac{\mathsf{Cov}\,(Y_i, Y_j)}{\bar{p}(1-\bar{p})} \\
&= \frac{\mathsf{E}\,[Y_i Y_j] - \mathsf{E}\,Y_i \mathsf{E}\,Y_j}{\bar{p}(1-\bar{p})} \\
&= \frac{\mathsf{E}\left[\mathsf{P}[Y_i = 1, Y_j = 1|P]\right] - \bar{p}^2}{\bar{p}(1-\bar{p})} \\
&= \frac{\int_0^1 p^2\,\mathrm{d}F(p) - \bar{p}^2}{\bar{p}(1-\bar{p})} \\
&= \frac{\mathsf{Var}\,(P)}{\bar{p}(1-\bar{p})}.
\end{aligned} \tag{4.19}$$

From (4.19) a few interesting consequences follows. First of all, it implies that the correlation between default events vanishes only when the variance of the random parameter P goes to 0, that is, F is the Dirac measure with mass in $\bar{p}$. The second immediate consequence of (4.19) is, that since the variance is always a positive number, it is impossible to capture the negative correlation structure by this model.

Now let us assume that we use Merton's model in a Brownian motion setting, and we are interested if the firm's asset value is at maturity below the face value of the debt or it is not. The default probability is

$$\bar{p} = \mathsf{P}[V_T < K] = \mathsf{P}\left[X_T < \log\frac{K}{V_0}\right].$$

We assume that X_T is normally distributed with mean value μT and variance $\sigma^2 T$. Then the default probability can be rewritten as

$$\bar{p} = \mathsf{P}\left[\frac{X_T - \mu T}{\sigma\sqrt{T}} < \frac{\log\frac{K}{V_0} - \mu T}{\sigma\sqrt{T}}\right] = \mathsf{P}\left[\tilde{X} < \tilde{K}\right],$$

where $\tilde{X}$ is a standard normal random variable with the cumulative distribution function Φ and $\tilde{K}$ is the normalized face value of the debt. Then it follows

$$\tilde{K} = \Phi^{-1}(\bar{p}), \tag{4.20}$$

because in the homogeneous portfolio the default probability of every debtor is the same.

Let us assume that the risk driver $\tilde{X}$ follows a one–factor model, i.e.,

$$\tilde{X} = \sqrt{\varrho}Z + \sqrt{1-\varrho}\,\varepsilon, \tag{4.21}$$

where $Z \sim N(0,1)$ is a systematic risk factor and $\varepsilon \sim N(0,1)$ is an idiosyncratic risk independent of Z, and $\sqrt{\varrho}$ is the load describing how much the debtor is sensitive to the systematic risk factor. Let

$$p(Z) = \mathsf{P}[Y = 1|Z] = \mathsf{E}\,[Y|Z],$$

be the default probability of debtor conditioned by the realization of the systematic risk factor Z. Then using (4.20) and (4.21) it follows

$$\begin{aligned} p(Z) &= \mathsf{P}\left[\sqrt{\varrho}Z + \sqrt{1-\varrho}\varepsilon < \tilde{K}\right] \\ &= \mathsf{P}\left[\varepsilon < \frac{\Phi^{-1}(\bar{p}) - \sqrt{\varrho}Z}{\sqrt{1-\varrho}}\right] \\ &= \Phi\left(\frac{\Phi^{-1}(\bar{p}) - \sqrt{\varrho}Z}{\sqrt{1-\varrho}}\right). \end{aligned} \tag{4.22}$$

The conditioned default probability $p(Z)$ is then a random variable with values in [0,1] and expected value $\mathsf{E}\,p(Z) = \bar{p}$. Now we can calculate the joint probability distribution of default events Y_i as

$$\mathsf{P}[Y_1 = y_1, \ldots, Y_n = y_n] = \int_{-\infty}^{\infty} p(z)^k \big(1 - p(z)\big)^{n-k} \varphi(z)\,\mathrm{d}z,$$

where $k = \sum_{i=1}^{n} y_i$, $p(z)$ is given by (4.22) and φ is the standard normal density function.

Heterogeneous Portfolios Let us try to extend the result of the homogeneous portfolio to the heterogeneous one. Let the default probability of debtor i be p_i and let the risk driver $X_T^{(i)}$ be given by the multi–factor model

$$X_T^{(i)} = \alpha^{(i)} \cdot \boldsymbol{Z} + \beta^{(i)}\varepsilon^{(i)},$$

where $\boldsymbol{Z}$ is a random vector normally distributed with a zero mean value and a covariance matrix Σ, and $\varepsilon^{(i)}$ are independent standard random variables, $\boldsymbol{\alpha^{(i)}}$ is a vector of constant factor loads, and β_i is also constant. Then the conditioned default probability is

$$p_i(\boldsymbol{Z}) = \Phi\left(\frac{\Phi^{-1}(p_i) - \alpha^{(i)} \cdot \boldsymbol{Z}}{\beta^{(i)}}\right),$$

and the joint default probability is

$$\mathsf{P}[Y_1 = y_1, \ldots, Y_n = y_n] = \int_{\mathbb{R}^n} \prod_{i=1}^{n} p_i^{y_i}(\boldsymbol{z})\big(1 - p_i(\boldsymbol{z})\big)^{1-y_i} \varphi(\boldsymbol{z}; \Sigma)\, \mathrm{d}\boldsymbol{z},$$

where $\varphi(\boldsymbol{z}; \Sigma)$ is the n-dimensional normal density function with zero mean and covariance matrix Σ. We see that the joint probability distribution is modeled the same as in the case of the homogeneous portfolio by the Gaussian copula (see Section B).

In the case of reduced–form models we will end up also with some copula (sometimes called *survival copulas*). For more, see Giesecke (2004, Section 3.6).

4.6.4 Monte Carlo Simulation

Let us assume that we have chosen some model dynamics, calibrate its parameters as well as the correlation structure and recoveries. Now we are interested in the probability distribution of the aggregate loss of the whole portfolio in some given horizon (for example one year). In some simple models we have seen that we can get a closed formula, but by adding more realistic features to the model and allowing more randomness and a dependency between different random drivers (like dependent recoveries and default events) we get a more and more complex problem, in which it is very hard to get a closed formula. In that case the simulation seems to be the only possibility.

Every run of the simulation gives us a realization of the random credit loss. We should perform sufficiently many runs of the simulation to get an image of the aggregate loss distribution. The complex simulation can take many days of the computation. Therefore, the number of runs of the simulation must be the compromise between the time consumption of the simulation and the precision of the resulting probability distribution.

To lower time requirements of the simulation some approximation methods can be used.

Approximations of Aggregate Loss Distribution

After n runs of the simulation we get n realizations of the random credit loss variable. We can assume that the loss distribution is very close or even belongs to some well–known parametric distribution family. Then we think about n realizations of the simulation as about a random sample from the distribution F. Parameters of the distribution F can be estimated using the moment method or the maximum likelihood method.

An often used distribution is the gamma distribution with density function

$$f(x) = \frac{a^b}{\Gamma(b)} e^{-ax} x^{b-1}, \qquad x > 0,$$

which is right–skewed and easily parameterized. If we are interested in modeling the number of defaults, i.e., modeling of the random variable

$$L = \sum_{i=1}^{m} \mathbf{1}_{\{\tau_i \le 1\}},$$

an often used distribution is the negative Binomial distribution. The integer–valued negative Binomial distribution can be interpreted as the gamma mixture of Poisson distributions which is consistent with the idea of reduced–form models where a default is the first jump of a Poisson process. If every defaulter has a Poisson process as its own default process and parameter λ is the random variable from the gamma distribution, then the random variable L has a

negative Binomial distribution. Confidence intervals about parameters of the distribution can be estimated through the *bootstrap*. Another possibility for the aggregate loss distribution estimation is to use non–parametric kernel estimators.

5
Real Data Study

In this chapter we illustrate the performance of estimation methods for partially observed time–changed data from Chapter 3 on real data.

5.1 Data Analysis and Estimations

We use the data set from Standard&Poor's (2006), where yearly transition matrices for the period 1981–2005 are given as well as the number of firms starting within a particular rating. Let us denote the transition matrices by $\mathbf{P}_1, \dots, \mathbf{P}_{25}$. We assume that the transition matrices $\mathbf{P}_1, \dots, \mathbf{P}_{25}$ were computed using the cohort method, i.e., for fixed k the ij-th element p_{ij}^k of matrix $\mathbf{P}_k$ was computed as

$$p_{ij}^k = \frac{n_{ij}^k}{n_{i.}^k},$$

where n_{ij}^k is the number of companies which started the k-th year in rating i and ended in rating j, and $n_{i.}^k$ is the number of companies started in rating i. Since we are given $\mathbf{P}_k$ and $n_{i.}^k$, we can compute elements of the frequency matrix $\mathbf{N}_k$ as

$$n_{ij}^k = p_{ij}^k n_{i.}^k. \tag{5.1}$$

This is of course valid only under the assumption that transition matrices were really computed by the cohort method.[27] The rating scale of S&P consists of 8 elements

$$\{AAA, AA, A, BBB, BB, B, CCC, D\},$$

and one special rating NR (not rated). The best rating is AAA and D is a default. S&P uses for their own purposes a more finer rating scale with much more ratings, but in publicly available data they are grouped into the 8 ratings above. Since we do not know what happened with companies that migrate to NR, we simply ignored these transitions, that is, we took just the first 8 columns of the frequency matrices $\mathbf{N}_1, \dots, \mathbf{N}_{25}$. The data after cleaning can be found in Appendix C.

Now let us recall what we want to show on real data. We assume that the system (ratings of hundreds of companies around the world monitored by S&P) follows a time–changed continuous–time Markov chain. We assume the TEPG setting (Section 3.2.2). After we will estimate the time evolution we will try to fit it by gama distribution. The time evolution enters into the system as parameters $\boldsymbol{t} = (t_1, \dots, t_{25})$. In other words we assume that the generator matrix in the k-th year is $t_k\mathbf{Q}$, where $\mathbf{Q}$ is same for every $k = 1, \dots, 25$. For the uniqueness of the solution we add a condition

$$\sum_{k=1}^{25} t_k = 25,$$

[27] Frequency data which we get by (5.1) was always very near to the integer value (6.99, 81.01,...), from which we can speculate that really the cohort method was used.

which forces same total business time and calendar time. We apply the CO method (Section 3.3.4) and the EM method (Section 3.3.2). Both methods give us the estimate $\boldsymbol{t}$ and $\mathbf{Q}$, but they solve a quite different problem. The CO method gives us the generator matrix which generate the transition probability matrix as close as possible to the empirical transition matrix. The EM method rather estimates "the most likely" generator matrix. All methods from Section 3.3 solve the first or the second problem. Naturally, the CO method gives the most accurate result for the first problem. The second problem can be also solved using MCMC method which according to simulation studies (for example Inamura (2006)) gives very similar results as the EM method. Hence we choose the CO and EM method for this demonstration.

Since we don't know the true generator matrix and the time evolution, it is hard to compare the performance of these two methods between each other. Hence in this section we compare mainly the fit of the time–unchanged CO method and the CO method with time change. In Section 5.2 we simulate the run of the system and compare the performance of wrongly used classical ML estimator and the CO and EM methods both time–unchanged and time–changed. Further, if we speak about distance of two matrices, the distance in Euclidean norm is meant by that. Let us introduce some notation used in this chapter:

COTC time–changed CO method

EMTC time–changed EM method

$\mathbf{Q}^{COTC}$ generator matrix estimated using COTC method

$\mathbf{P}^{COTC}$ transition probability matrix generated by $\mathbf{Q}^{COTC}$, i.e., $\mathbf{P}^{COTC} = \exp(\mathbf{Q}^{COTC})$

$\left\|\mathbf{P}^{COTC} - \mathbf{P}\right\|$ Euclidean distance of matrices $\mathbf{P}^{COTC}$ and $\mathbf{P}$

$\boldsymbol{DP}$ default probabilities (last column of matrix $\mathbf{P}$)

$\boldsymbol{DP}^{COTC}$ default probabilities (last column of matrix $\mathbf{P}^{COTC}$) estimated using COTC method

Similar for EMTC.

On an accompanied CD the implementation of all methods derived in this work can be found. All methods are implemented in Matlab.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	−0.0903	0.0873	0.0018	0.0006	0.0006	0.0000	0.0000	0.0000
AA	0.0067	−0.0994	0.0871	0.0039	0.0003	0.0011	0.0002	0.0000
A	0.0005	0.0217	−0.0920	0.0646	0.0032	0.0014	0.0003	0.0003
BBB	0.0002	0.0014	0.0453	−0.1101	0.0524	0.0067	0.0020	0.0022
BB	0.0004	0.0004	0.0015	0.0670	−0.1859	0.0971	0.0110	0.0085
B	0.0000	0.0006	0.0022	0.0013	0.0752	−0.2046	0.0707	0.0545
CCC	0.0000	0.0000	0.0041	0.0058	0.0135	0.1870	−0.6215	0.4111
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5.1. Generator estimation using CO method (without time change).

In Tables 5.1 - 5.4 you can see generator estimators using CO and EM method both time–unchanged and time–changed. Time evolution estimators in the case of time–changed CO and EM method can be seen in Table 5.6.

The main purpose of these modeling is to estimate default probabilities (last column of transition matrices). Figure 5.1 shows us the Euclidean distance of the theoretical default probabilities $\boldsymbol{DP}_k^{CO}$ and $\boldsymbol{DP}_k^{COTC}$ from the empirical default probabilities $\boldsymbol{DP}_k$. Figure 5.1 also shows the time evolution estimated using COTC. A better fit is almost in every year for $\boldsymbol{DP}^{COTC}$ matrix. Detailed distances and percentage improvements are in Table 5.5. The overall improvement in fitting is 35 %. Notice that we are comparing the CO and COTC method. In practice one often uses the maximum likelihood (ML) method. This wrong usage

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	−0.0903	0.0873	0.0018	0.0007	0.0006	0.0000	0.0000	0.0000
AA	0.0067	−0.0994	0.0871	0.0039	0.0004	0.0011	0.0002	0.0000
A	0.0005	0.0217	−0.0920	0.0646	0.0032	0.0014	0.0003	0.0003
BBB	0.0002	0.0014	0.0453	−0.1101	0.0524	0.0067	0.0020	0.0022
BB	0.0004	0.0004	0.0015	0.0670	−0.1859	0.0971	0.0110	0.0085
B	0.0000	0.0006	0.0022	0.0013	0.0752	−0.2046	0.0707	0.0545
CCC	0.0000	0.0000	0.0040	0.0059	0.0135	0.1870	−0.6215	0.4111
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5.2. Generator estimation using EM method (without time change).

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	−0.0830	0.0798	0.0020	0.0006	0.0006	0.0000	0.0000	0.0000
AA	0.0062	−0.0993	0.0885	0.0030	0.0006	0.0008	0.0002	0.0000
A	0.0002	0.0191	−0.0875	0.0632	0.0033	0.0013	0.0002	0.0003
BBB	0.0002	0.0016	0.0451	−0.1120	0.0543	0.0069	0.0016	0.0023
BB	0.0003	0.0006	0.0010	0.0715	−0.1826	0.0879	0.0121	0.0092
B	0.0000	0.0008	0.0025	0.0010	0.0733	−0.1953	0.0606	0.0571
CCC	0.0000	0.0000	0.0026	0.0081	0.0116	0.1597	−0.5541	0.3722
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5.3. Generator estimation using COTC method.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	−0.0904	0.0874	0.0017	0.0007	0.0006	0.0000	0.0000	0.0000
AA	0.0067	−0.0997	0.0877	0.0036	0.0003	0.0011	0.0002	0.0000
A	0.0004	0.0221	−0.0935	0.0659	0.0031	0.0015	0.0003	0.0003
BBB	0.0002	0.0014	0.0463	−0.1127	0.0541	0.0066	0.0020	0.0022
BB	0.0004	0.0004	0.0014	0.0686	−0.1906	0.1005	0.0110	0.0083
B	0.0000	0.0006	0.0023	0.0013	0.0768	−0.2088	0.0727	0.0552
CCC	0.0000	0.0000	0.0038	0.0056	0.0125	0.1838	−0.6101	0.4043
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5.4. Generator estimation using EMTC method.

(due to the partial observations) of the ML method can give a much worse fit than the CO method.

In Figure 5.2 we can see the time evolution estimation from both methods (COTC and EMTC). They are quite similar, though COTC estimator seems a slightly more volatile. We can clearly see three periods where the time evolution was significantly higher (especially in COTC method) than in remaining years. The first period are years 1990–1993 which can be connected to the U.S. recession in years 1990–1991. The second period is the year 1998 in which the Russian crisis was and the last period is 2001–2002 when the Internet bubble burst.

These observations prove our suggestion that if the economy is in recession the time evolution is quicker and companies default more often. On the contrary when there is an economic expansion the time evolution slows down and companies default less often.

We would like to find some evidence of the connection between the time evolution and the state of economy. As indicators of the state of the economy we took the federal fund target rate (FFTR)[28], the American GDP growth and growth of index S&P500[29]. Though the

[28] The interest rate announced by the American FED. It is a major tool for regulation of the economy.

[29] By index growth for year 1981 (and similar for other years) we mean index value in the beginning of the year 1981 divided by the starting value in year 1980.

migration data are about companies all around the world we took all three indicators from the American economy, because the American economy has great influence on the world economy. We had FFTR data from the Federal Reserve Bank of New York (2008), GDP data from the Economic Research Division of Federal Reserve Bank of St. Louis (2008) and S&P500 index quotes from website Finance Yahoo[30]. All three of them are in Table 5.6.

We have computed the correlation of estimated time evolution and FFRT, resp. GDP growth, resp. S&P500 index growth. In all cases a high negative correlation was the result. Precise values are in Table 5.7 From this correlation values COTC method seems to be a slightly "better" method for time evolution estimation. However further research can try to explain and predict the time evolution using some econometric models, where EM time evolution can be explained much better. Of course for that we need better and more data.

5.1.1 Fitting Time Evolution by Distribution

Let us try to model the estimated time evolution $\boldsymbol{t}_{COTC}$ and $\boldsymbol{t}_{EMTC}$ as a Lévy subordinator. We choose the Gamma process for our model. Using the maximum likelihood we estimate parameters of one–year gamma distributed increments with the density function

$$f(x) = \frac{a^b}{\Gamma(b)} e^{-ax} x^{b-1}, \qquad x > 0,$$

[30] `http://finance.yahoo.com/`

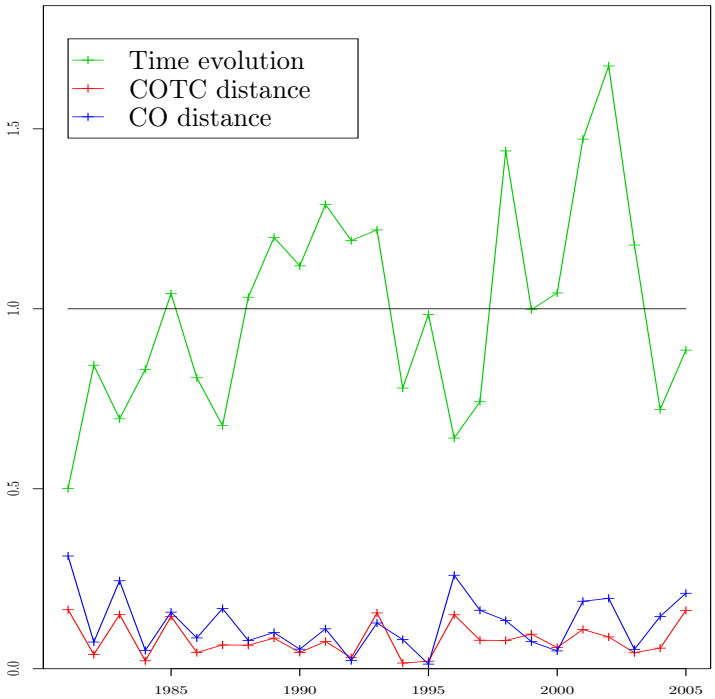


Figure 5.1. COTC time evolution and default probability fit.

Estimated parameters are

$$a_{COTC} = b_{COTC} = 12.5095,$$
$$a_{EMTC} = b_{EMTC} = 27.209.$$

One can be wondering why in both cases $a = b$. The explanation is very easy since mean value of gamma distribution is $\frac{a}{b}$ and during estimation we forced total business time equal to total calendar time which imply sample mean equal 1 and hence $a = b$ is natural consequence.

We would like to know how good our hypothesis is that time evolves according to the Gamma process. We tested using the goodness–of–fit test (see, for example, Lehmann and Romano (2005, Section 14.3)) whether we can reject the hypothesis of the Gamma distribution. In both cases we don't reject hypothesis of Gamma process with p-values

$$p\text{-value}^{COTC} = 0.819,$$
$$p\text{-value}^{EMTC} = 0.247,$$

we did not reject the Gamma distribution. In Figures 5.3 and 5.4 we can see the histogram of our observations and the theoretical density function.

5.2 EM and CO Method Simulation Comparison

In the previous section we used time–changed CO and EM method. We would like to be able to compare the performance of these methods. Hence we did a simulation where the system

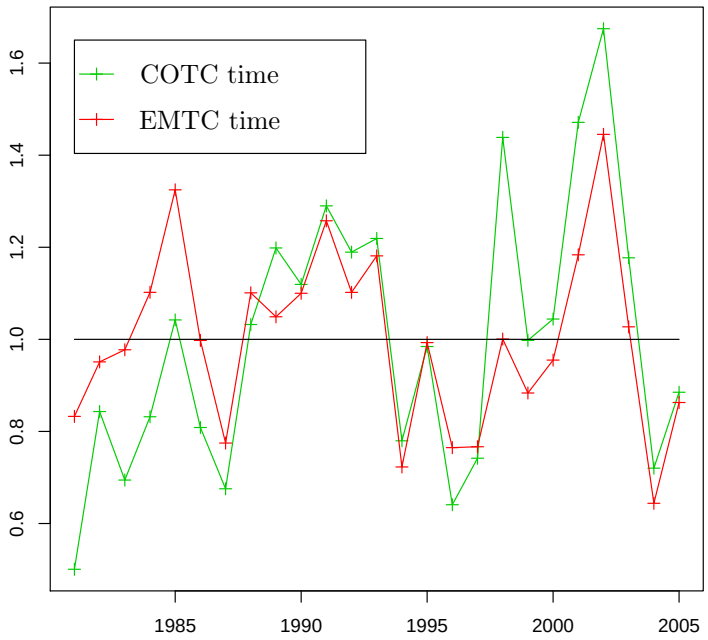


Figure 5.2. COTC and EMTC time evolution estimates.

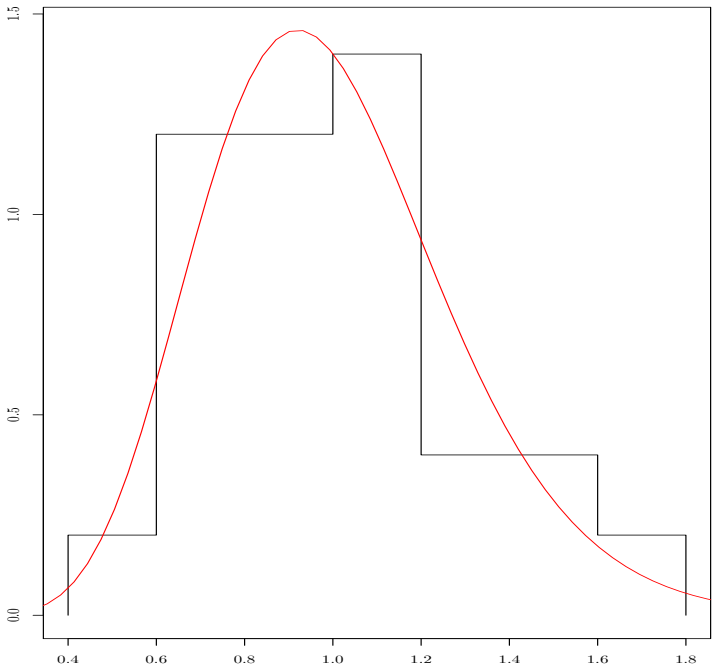


Figure 5.3. Time evolution histogram and theoretical density for COTC.

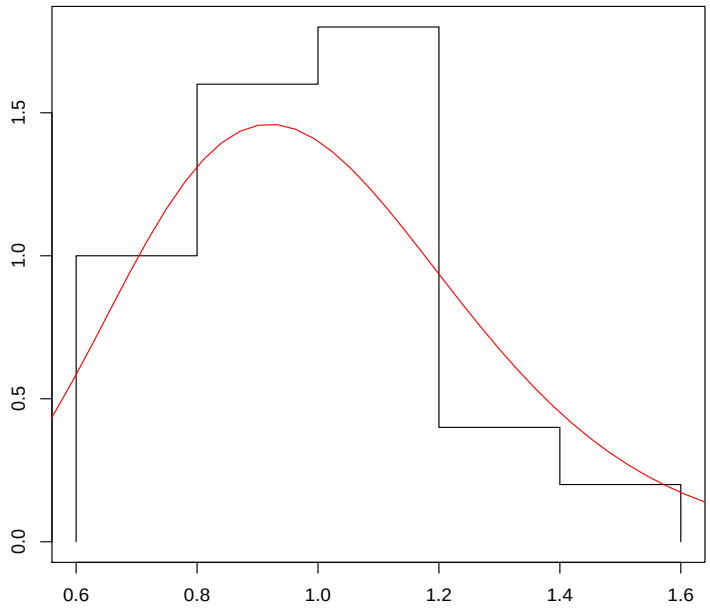


Figure 5.4. Time evolution histogram and theoretical density for EMTC.

year	$\|\|\boldsymbol{DP} - \boldsymbol{DP}^{CO}\|\|$	$\|\|\boldsymbol{DP} - \boldsymbol{DP}^{COTC}\|\|$	Improvement
1981	0.313	0.164	−47.64 %
1982	0.074	0.040	−46.63 %
1983	0.244	0.150	−38.49 %
1984	0.051	0.022	−56.79 %
1985	0.157	0.145	−7.55 %
1986	0.086	0.045	−47.54 %
1987	0.167	0.066	−60.44 %
1988	0.078	0.065	−16.52 %
1989	0.101	0.085	−15.61 %
1990	0.054	0.045	−16.54 %
1991	0.111	0.076	−31.61 %
1992	0.023	0.031	36.08 %
1993	0.127	0.155	22.05 %
1994	0.081	0.016	−80.53 %
1995	0.013	0.021	62.43 %
1996	0.260	0.150	−42.25 %
1997	0.162	0.079	−51.30 %
1998	0.134	0.078	−41.49 %
1999	0.075	0.096	27.72 %
2000	0.050	0.059	18.83 %
2001	0.187	0.109	−41.85 %
2002	0.196	0.088	−54.85 %
2003	0.054	0.044	−17.50 %
2004	0.145	0.057	−60.65 %
2005	0.210	0.162	−22.72 %
1981–2005	3.153	2.049	−35.00 %

Table 5.5. Comparison of the default probabilities fit by CO and COTC method.

dynamics follows a time–changed continuous–time Markov chain. As a generator matrix and time evolution we took estimators from the previous section. We simulated system run during 25 years. We perform whole simulation 1000 times. For every run we estimate the generator matrix using classical ML estimator, time–unchanged CO and EM method and time–changed CO and EM method. For every run and every year we computed the distance of estimated transition matrices from the correct ones and also distance of default probabilities of from the correct ones. Mean values which we get are in Table 5.8. We clearly see that using time–changed method we get more than twice more accurate results than in the case of using classical ML estimation and 50% improvement in comparison with time–unchanged methods. We can also conclude that COTC method outperform EMTC method and hence we can revise results in previous section in a favor of COTC method, which gave us "nicer" results.

year	t_{COTC}	t_{EMTC}	FFTR	GDP growth	S&P500 growth
1981	0.501	0.833	12.00	12.15	1.135
1982	0.843	0.951	8.50	4.05	0.929
1983	0.694	0.977	9.50	8.65	1.207
1984	0.832	1.102	8.25	11.21	1.125
1985	1.042	1.325	7.75	7.30	1.099
1986	0.809	0.998	6.00	5.75	1.179
1987	0.675	0.775	6.75	6.20	1.294
1988	1.032	1.101	9.00	7.69	0.938
1989	1.199	1.049	8.25	7.46	1.157
1990	1.119	1.100	7.00	5.81	1.106
1991	1.290	1.258	4.00	3.32	1.045
1992	1.189	1.102	3.00	5.70	1.189
1993	1.219	1.181	3.00	5.04	1.073
1994	0.780	0.723	5.50	6.23	1.098
1995	0.984	0.993	5.50	4.60	0.977
1996	0.641	0.765	5.25	5.67	1.352
1997	0.742	0.767	5.50	6.24	1.236
1998	1.439	1.000	4.75	5.33	1.247
1999	0.998	0.883	5.50	5.96	1.305
2000	1.044	0.955	6.50	5.92	1.090
2001	1.471	1.184	1.75	3.17	0.980
2002	1.675	1.445	1.25	3.37	0.827
2003	1.177	1.027	1.00	4.69	0.757
2004	0.720	0.644	2.25	6.86	1.322
2005	0.885	0.863	4.25	6.35	1.044

Table 5.6. Time evolution, FFTR, GDP growth and index S&P growth data.

Corr	FFTR	GDP growth	S&P500 growth
T_{COTC}	-0.574	-0.608	-0.498
T_{EMTC}	-0.187	-0.285	-0.570

Table 5.7. Computed correlations of estimated evolution and economy indicators.

$(\cdot)$	classical ML	CO	EM	COTC	EMTC
average $\|\|\mathbf{P} - \mathbf{P}^{(\cdot)}\|\|$	4.499	3.573	3.573	2.329	2.443
average $\|\|\boldsymbol{DP} - \boldsymbol{DP}^{(\cdot)}\|\|$	2.054	1.618	1.618	0.927	1.106

Table 5.8. Average transition probabilities and default probabilities distance comparison of classical ML, CO, EM, COTC and EMTC method.

6
Summary

In the presented work we described commonly used credit risk models. An important part of credit risk models is modeling of the credit ratings migration, which is in practice often done using the time homogeneous continuous–time Markov chain. However, many case studies showed the contradict with this time homogeneity assumption. If we want to introduce the time inhomogeneity, we often do not have a clue about the inhomogeneity structure and even if we have, it is hard to estimate all necessary parameters of the model since the inhomogeneity structure often introduces too many new parameters into the model.

In this work we came up with an idea, that **the credit ratings follow a time homogeneous continuous–time Markov chain but with respect to different time**, which we called a business time. The business time runs faster when the economy is in recession, which means clients migrate to lower ratings faster (and default faster) and contrary in a time of economic growth, the business time runs slower (less defaults). This simple extension allows us to introduce easily handled inhomogeneity into the model — only through the time change.

In Chapter 3 we extended the classical theory of Markov chains using the time change. Probably the most important thing for practice modeling is parameter estimation. A lot of practitioners make a mistake when they use ML estimator. ML estimator assume fully observed data, but in the practice often only annual or semiannual data are available. In Section 2.3 is described how to proceed if we have a partially observed data. In Chapter 3 we derived a ML estimators as well as estimators in the case of partially observed data. Further, we demonstrated the performance of our methods in the last chapter on a real data.

For further research remains modeling of the business time dependency on macroeconomic variables. If we could predict the business time, we would get more precise estimations of the credit migration process. Other interesting area where business time could be applied is the recovery rates. The majority of models assume that recovery rates are independent of default probabilities despite the fact that many case studies show that they are highly positively correlated. An extension of these models where the default probability and recovery rates would be independent conditionally on the realization of business time, would provide an easy way how to deal with the problem.

The provided extension of a continuous–time Markov chain to the stochastic time in this thesis is very easy doable and its performance is at least as good as the original model. In our data we showed a 35 % better fit in default probabilities. We think it is not a negligible improvement. If we could find the dependency of the time evolution on some observable variables using some econometric models, we could predict more precisely future transition matrices.

Appendix

A Lévy Processes

Classical references for Lévy process theory are Bertoin (1996) and Sato (1999). An overview of the most common Lévy processes with references to original papers can be found in Schoutens (2003, Chapter 5).

A.1 Lévy Processes

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathsf{P})$ be the probability space equipped by filtration $\mathcal{F}_t, t \geq 0$ that satisfies the "usual hypothesis", i.e., is right–continuous and for every $t \geq 0$ the σ-algebra $\mathcal{F}_t$ contains all null sets of the σ-algebra $\mathcal{F}$. We will describe properties for a one–dimensional Lévy process. For multi–dimensional properties see the literature above.

A stochastic cadlag[31] process X_t adapted to filtration $\mathcal{F}_t$ with values in $\mathbb{R}$ is a Lévy process if and only if it has following properties:

(L1) $X_0 = 0$ a.s.

(L2) For every sequence of times $t_0 < t_1 < \ldots < t_n$ the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.

(L3) The process X_t has stationary increments, i.e., for every $t > 0$ and $s > 0$

$$\mathcal{L}(X_{t+s} - X_t) = \mathcal{L}(X_s),$$

where $\mathcal{L}(X)$ denotes the law of random variable X.

(L4) The process X_t is continuous in probability, i.e., for every $\varepsilon > 0$ holds

$$\lim_{h\to 0} \mathsf{P}[|X_{t+h} - X_t| > \varepsilon] = 0.$$

The cadlag property is not necessary in the definition of the Lévy process, but as we will see later, the Lévy process is a semi–martingale[32] and every semi–martingale admits the unique modification that is cadlag. Therefore we have included the cadlag property in the definition.

We say that the distribution μ is *infinitely divisible* if for every $n \in \mathbb{N}$ there exist n i.i.d. random variables $Y_1, \ldots, Y_n$ such that the random variable $Y = Y_1 + \ldots + Y_n$ has the probability distribution μ. If μ_n is the probability distribution of random variables $Y_1, \ldots, Y_n$ then μ is the n-th convolution of μ_n, i.e. $Y = Y_1 * \ldots * Y_n$.

There exists a close connection between infinitely divisible distributions and Lévy processes. In fact for every Lévy process X_t there exists an infinitely divisible distribution μ such that $\mathcal{L}(X_1) = \mu$. Conversely for any infinitely divisible distribution μ there exists a Lévy process with $\mathcal{L}(X_1) = \mu$. Indeed, for every $n \in \mathbb{N}$ we can write X_1 as

$$X_1 = X_{\frac{1}{n}} + \left(X_{\frac{2}{n}} - X_{\frac{1}{n}}\right) + \ldots + \left(X_{\frac{n}{n}} - X_{\frac{n-1}{n}}\right),$$

[31] The cadlag process is a process with right–continuous sample path which admits a left limit.

[32] We say that a process X_t is a semi–martingale if and only if

$$X_t = X_0 + M_t + C_t,$$

where M_t is a local martingale and C_t is a cadlag adapted process of a locally bounded variation.

and since X_t has independent stationary increments it follows that

$$\mathcal{L}(X_1) = \mathcal{L}\left(X_{\frac{1}{n}} + \left(X_{\frac{2}{n}} - X_{\frac{1}{n}}\right) + \ldots + \left(X_{\frac{n}{n}} - X_{\frac{n-1}{n}}\right)\right) = \mathcal{L}\left(X_{\frac{1}{n}}\right) * \ldots * \mathcal{L}\left(X_{\frac{1}{n}}\right),$$

hence the distribution of X_1 is the n-th convolution of the distribution of $X_{\frac{1}{n}}$ and therefore the distribution of X_1 is infinitely divisible. The converse implication can be proved by the construction of the process using the Kolmogorov theorem and verifying that the constructed process is a Lévy process. Details of the construction can be found in Sato (1999, Theorem 7.10).

Let us define the characteristic function of X_t as

$$\phi_{X_t}(z) = \mathsf{E}\left[\mathrm{e}^{izX_t}\right], \qquad z \in \mathbb{R}.$$

For the characteristic function $\phi_{X_{t+s}}(z)$, where $t, s > 0$ it holds

$$\begin{aligned}
\phi_{X_{t+s}}(z) &= \mathsf{E}\left[\mathrm{e}^{izX_{t+s}}\right] \\
&= \mathsf{E}\left[\mathrm{e}^{iz(X_{t+s}-X_s+X_s)}\right] \\
&= \mathsf{E}\left[\mathrm{e}^{iz(X_{t+s}-X_s)}\right]\mathsf{E}\left[\mathrm{e}^{izX_s}\right] \\
&= \mathsf{E}\left[\mathrm{e}^{izX_t}\right]\mathsf{E}\left[\mathrm{e}^{izX_s}\right] \\
&= \phi_{X_t}(z)\phi_{X_s}(z), \qquad ,
\end{aligned} \tag{1}$$

since the Lévy process has independent stationary increments. Now let $t \in \mathbb{Q}^+$, then there exists $m, n \in \mathbb{N}$ such that $t = \frac{m}{n}$ and from (1) it follows

$$\phi_{X_m}(z) = \left(\phi_{X_{\frac{m}{n}}}(z)\right)^n,$$

and

$$\phi_{X_t}(z) = \phi_{X_{\frac{m}{n}}}(z) = \left(\phi_{X_m}(z)\right)^{\frac{1}{n}} = \left(\phi_{X_1}(z)\right)^{\frac{m}{n}}.$$

Then from the continuity in probability it follows that for every $t \geq 0$

$$\phi_{X_t}(z) = \left(\phi_{X_1}(z)\right)^t = \mathrm{e}^{t\psi(z)},$$

where the exponential form is implied by property (1) and ψ is called the *characteristic exponent*.

A.2 Examples of Lévy Processes

We mention just few examples of Lévy processes. For more see (Schoutens, 2003, Chapter 5).

Brownian motion

A *standard Brownian motion* W_t is a Lévy process such that

1. $W_t \sim N(0, t)$ for each $t \geq 0$.
2. W_t has continuous sample paths.

From property 1 it follows that the characteristic function of the Brownian motion is

$$\phi_{W_t}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{izx}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x = \exp\left(-\frac{1}{2}tz^2\right).$$

The Brownian motion is a very useful construction used in many different areas for modeling various phenomena. Let $b \in \mathbb{R}$ and $\sigma > 0$, then the process

$$B_t = bt + \sigma W_t,$$

is called the *Brownian motion with drift*. The process B_t is a Gaussian Lévy process with $B_t \sim N(bt, \sigma^2 t)$. Its characteristic function is given by

$$\phi_{B_t}(z) = \exp\left(ibtz - \frac{1}{2}t\sigma^2 z^2\right).$$

Poisson Process

We have already defined a Poisson process N_t in Section 2.1.2. The Poisson process is a Lévy process such that N_t has a Poisson distribution with parameter λt, i.e.,

$$\mathsf{P}[N_t = k] = \frac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda t}, \qquad \text{for } k = 0, 1, \ldots.$$

Therefore the characteristic function of the Poisson process is

$$\phi_{N_t}(z) = \sum_{k=0}^{\infty} \mathrm{e}^{izk}\frac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda t} = \exp\left(\lambda t(\mathrm{e}^{iz} - 1)\right).$$

Recall that the Poisson process is the only counting process with independent stationary increments.

From Section 2.1.2 we know that the Poisson process has the compensator λt. Therefore we can define *compensated Poisson process* $\tilde{N}_t$ as

$$\tilde{N}_t = N_t - \lambda t,$$

which is a martingale.

Compound Poisson process

A Poisson process is a process with positive jumps of size 1. If we generalize a Poisson process and allow a random size of jumps we get the *compound Poisson process*. More precisely, let N_t be a Poisson process and $Y_1, Y_2, \ldots$ i.i.d. random variables having probability distribution F independent with Poisson process N_t. Then the process

$$X_t = \sum_{i=1}^{N_t} Y_i, \qquad t \geq 0,$$

is the compound Poisson process.[33] The compound Poisson process jumps at the same times as the original Poisson process N_t, but has random size of jumps. The characteristic function

[33] We assume $\sum_1^0 \cdot = 0$.

of the compound Poisson process is

$$\begin{aligned}
\phi_{X_t}(z) &= \mathsf{E}\, \mathrm{e}^{izX_t} \\
&= \mathsf{E}\, \left[\mathsf{E}\, [\mathrm{e}^{iz\sum_{i=1}^{N_t} Y_i} | N_t] \right] \\
&= \mathsf{E}\, \left[\left(\mathsf{E}\, [\mathrm{e}^{izY_1}] \right)^{N_t} \right] \\
&= \sum_{k=0}^{\infty} \left(\mathsf{E}\, [\mathrm{e}^{izY_1}] \right)^k \frac{(\lambda t)^k}{k!} \mathrm{e}^{-\lambda t} \\
&= \exp\left(t \int (\mathrm{e}^{izy} - 1) \lambda \, \mathrm{d}F(y) \right).
\end{aligned}$$

α-stable distribution

An important subclass of infinitely divisible distributions are so called *stable distributions*.

We say that distribution μ is stable if and only if for all $n > 1$ there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$\mathcal{L}(X_1 + X_2 + \ldots + X_n) = \mathcal{L}(c_n X + d_n),$$

where $X, X_1, X_2, \ldots, X_n$ are i.i.d. random variables with distribution μ. Distribution μ is called *strictly stable* if and only if $d_n = 0$ for all n. It can be proved that the only possibility of choosing c_n is

$$c_n = n^{\frac{1}{\alpha}},$$

where $\alpha \in (0, 2]$ is called the *index of stability*. If X has index of stability α we say that X is α-stable.

The random variable X is stable if and only if

$$X = aY + b,$$

where $a > 0$, $b \in \mathbb{R}$ and Y is a random variable with the characteristic function

$$\mathsf{E}\, \mathrm{e}^{izY} = \begin{cases} \exp\left(-|z|^\alpha \left(1 - i\beta (\operatorname{sgn} z) \tan \frac{\pi\alpha}{2} \right) \right), & \alpha \neq 1, \\ \exp\left(-|z| \left(1 - i\beta (\operatorname{sgn} z)(-\frac{\pi}{2} \log |z|) \right) \right), & \alpha = 1, \end{cases}$$

where $-1 \leq \beta \leq 1$. For a proof see Sato (1999, Theorem 14.15). If $b = 0$ and $\beta = 0$ then X is symmetric around zero and the characteristic function of $X = aZ$ is simply

$$\mathsf{E}\, \mathrm{e}^{izaY} = \exp(-a^\alpha |z|^\alpha).$$

Some special cases of α-stable distributions are:

Normal distribution $N(\mu, \sigma^2)$

$$\begin{aligned} \alpha &= 2, & a &= \frac{\sigma}{\sqrt{2}}, \\ \beta &= 0, & b &= \mu. \end{aligned}$$

Cauchy distribution with parameters γ and δ

$$\begin{aligned} \alpha &= 1, & a &= \gamma, \\ \beta &= 0, & b &= \delta. \end{aligned}$$

Lévy distribution with parameters γ and δ

$$\alpha = \frac{1}{2}, \qquad a = \gamma,$$
$$\beta = 1, \qquad b = \delta.$$

One interesting thing is that the stable distribution has a finite second moment only for $\alpha = 2$ and finite first moment for $1 < \alpha \leq 2$.

A.3 Lévy–Ito Decomposition and Lévy–Khintchin Formula

First, we will mention a few properties of infinitely divisible distributions which are also properties of Lévy processes due to their straightforward connection. It is obvious that if μ and ν are independent infinitely divisible distributions then also the convolution $\mu * \nu$ is infinitely divisible. If μ_k is a sequence of infinity divisible distributions and $\mu_k \to \mu$ then also μ is infinitely divisible.

It can be even proved that every infinitely divisible distribution μ is a limit of the sequence of compounded Poisson processes with the distribution μ_n, even if μ is not a compounded Poisson process.

If every infinitely divisible distribution can be written as a limit of the sequence of compounded Poisson distributions, one can ask if there exists some general form for the characteristic function of a Lévy process. The answer is the Lévy–Khintchin formula. For better understanding of the formula we will present a few ideas that are behind it, but we will not prove it. The Lévy–Khintchin formula can be proved directly, see Sato (1999, Theorem 8.1), or by firstly proving the Lévy–Ito decomposition and then it follows as an easy corollary, see Cont and Tankov (2003, Theorem 3.1).

Let us define the *jump process* ΔX_t associated to the Lévy process X_t as

$$\Delta X = X_t - X_{t-}.$$

Since the Lévy process X_t is continuous in probability it follows that for fixed t, $\Delta X_t = 0$ a.s.. Some difficulties in Lévy processes modeling come from the fact that it is possible for any $t > 0$ to have[34]

$$\sum_{0 \leq s \leq t} |\Delta X_s| = \infty, \qquad \text{a.s.},$$

but we always have

$$\sum_{0 \leq s \leq t} |\Delta X_s|^2 < \infty, \qquad \text{a.s.}$$

Let $A \subset \mathbb{R}$ be any Borel set such that 0 does not belong to A, $0 \notin A$. Then we can define for $\omega \in \Omega$ a counting measure

$$N_t(A)(\omega) = \sum_{0 \leq s \leq t} \mathbf{1}_{\{\Delta X_s(\omega) \in A\}},$$

which assigns to the set A a number of jumps with size in A that has happened before time t. Since for any fixed $\omega \in \Omega$ and $t \geq 0$, $N_t(A)(\omega)$ is a counting measure on $\mathbb{R}$, $N_t(A)$ is a counting process.

[34] Continuous sum $\sum_{0 \leq s \leq t} |\Delta X_s|$ is nothing else then the sum of absolute values of all jumps up to time t.

It is easy to verify that $N_t(A) - N_s(A)$ is independent of $\mathcal{F}_s$. From stationarity of the Lévy process X_t it also follows that

$$\mathcal{L}\big(N_t(A) - N_s(A)\big) = \mathcal{L}\big(N_{t-s}(A)\big).$$

Let A be bounded from 0, i.e., 0 does not belong to closure of A, $0 \notin \bar{A}$, then $N_t(A) < \infty$ a.s.. From the fact that the counting process $N_t(A)$ has independent stationary increments, it follows that $N_t(A)$ is the Poisson process with intensity

$$\nu(A) = \mathsf{E}\, N_1(A),$$

where $\nu(A)$ is a σ-finite measure called the *Lévy measure*. Further, let us define the *compensated Poisson measure* $\tilde{N}_t(A)$ as

$$\tilde{N}_t(A) = N_t(A) - t\nu(A),$$

which is martingale, i.e., for $s < t$

$$\mathsf{E}\,[\tilde{N}_t(A)|\mathcal{F}_s] = \tilde{N}_s(A).$$

If A is not bounded from 0 then $\mathsf{E}\, N_t(A)$ and also $\mathsf{E}\, \tilde{N}_t(A)$ can be possibly infinite for any $t > 0$.

Now we would like to study the jump part of the Lévy process X_t, which should be intuitively equal to the sum of all jumps up to time t

$$\sum_{0 \le s \le t} \Delta X_s.$$

We must be careful here, since the Lévy process can have infinitely many small jumps in an arbitrary small time interval. That is reason why we divide jumps of the Lévy process X_t into two groups — jumps which are in absolute value smaller than some arbitrary constant $c > 0$ and the rest that are bigger or equal to c. It is irrelevant which c we choose since there might be infinitely many jumps smaller than any constant $c > 0$. The common choice for the constant c is 1. Then

$$\int_{|x| \ge 1} x N_t(\,\mathrm{d}x) = \sum_{0 \le s \le t} \Delta X_s \mathbf{1}_{\{|\Delta X_s| \ge 1\}},$$

is the sum of jumps of size bigger or equal to 1. This sum is a finite random variable since there are just finitely many jumps bigger then 1, but it does not need to have finite moments. On the other side

$$X_t - \int_{|x| \ge 1} x N_t(\,\mathrm{d}x), \qquad t \ge 0,$$

is a Lévy process with all moments finite. It can be shown that the integral

$$\int_{\varepsilon_n \le |x| < 1} x \tilde{N}_t(\,\mathrm{d}x),$$

converges almost surely and uniformly as $\varepsilon_n \downarrow 0$ (see Sato (1999, Lemma 20.6)). Hence the sum of compensated small jumps

$$\int_{|x| < 1} x \tilde{N}_t(\,\mathrm{d}x),$$

is a martingale. Finally it can be proved that

$$B_t = X_t - bt - \int_{|x|<1} x\tilde{N}_t(\,\mathrm{d}x) - \int_{|x|\geq 1} xN_t(\,\mathrm{d}x), \qquad t \geq 0,$$

with

$$b = \mathsf{E}\left(X_1 - \int_{|x|\geq 1} xN_1(\,\mathrm{d}x)\right),$$

is a martingale with continuous paths starting at 0 and a variance $t\sigma^2$, where $\sigma > 0$. From the Lévy theorem it follows that the process B_t is a Brownian motion.

We get the Lévy–Ito decomposition which says that if X_t is a Lévy process then there exists a Brownian motion B_t with drift $b \in \mathbb{R}$, variance $\sigma > 0$, and a random Poisson measure N_t such that

$$X_t = bt + B_t + \int_{|x|<1} x\tilde{N}_t(\,\mathrm{d}x) + \int_{|x|\geq 1} xN_t(\,\mathrm{d}x), \qquad t \geq 0,$$

where all terms are independent.

Recall that a process X_t is a semimartingale if and only if

$$X_t = X_0 + M_t + C_t,$$

where M_t is a local martingale and C_t is an adapted process of finite variation. We see that a Lévy process is a semimartingale where

$$M_t = B_t + \int_{|x|<1} x\tilde{N}_t(\,\mathrm{d}x),$$
$$C_t = bt + \int_{|x|\geq 1} xN_t(\,\mathrm{d}x).$$

Now if we compute the characteristic function of a general Lévy process using the Lévy–Ito decomposition we will get the Lévy–Khintchine formula

$$\begin{aligned}\mathsf{E}\,\mathrm{e}^{izX_t} &= \mathsf{E}\,\mathrm{e}^{t\psi(z)} \\ &= \exp\left(t\left(ibz - \frac{1}{2}\sigma^2 z^2 + \int_{\mathbb{R}} \left(\mathrm{e}^{izx} - 1 - izx\mathbf{1}_{\{|x|<1\}}\right)\nu(\,\mathrm{d}x)\right)\right), \qquad z \in \mathbb{R},\end{aligned}$$

where the Lévy measure ν satisfies the following two properties

$$\nu(\{0\}) = 0,$$
$$\int_{\mathbb{R}} (|x|^2 \wedge 1)\,\nu(\,\mathrm{d}x) < \infty.$$

The triplet (b, σ, ν) is called the characteristic triplet of the infinitely divisible distribution of X_1. This triplet uniquely defines the distribution of X_1. The characteristic triplet of the distribution of X_t is then $(tb, t\sigma, t\nu)$.

The jump part of the Lévy process X_t is of finite variation if and only if the Lévy measure ν satisfies

$$\int_{|x|<1} |x|\,\nu(\,\mathrm{d}x) < \infty.$$

In that case we can rewrite the characteristic exponent in the Lévy–Khintchin formula in the form

$$\psi(z) = ib'z - \frac{1}{2}\sigma^2 z^2 + \int_{\mathbb{R}} \left(\mathrm{e}^{izx} - 1\right) \nu(\,\mathrm{d}x),$$

where

$$b' = b - \int_{|x|<1} izx\ \nu(\,\mathrm{d}x).$$

We see that it is a convolution of a Brownian motion with drift b' and a compounded Poisson process with unit intensity and jumps size distribution ν.

B Copulas

Let the random variables X_i, $i = 1, \ldots, n$ from Ω to $(0, \infty)$ be some survival times. We are interested in how X_i depend on each other. More precisely we know $\mathsf{P}[X_1 \leq x_1], \ldots, \mathsf{P}[X_n \leq x_n]$ and we would like to model the joint probability $\mathsf{P}[X_1 \leq x_1, \ldots, X_n \leq x_n]$. The answer for that is to introduce *copulas*.

Let $\boldsymbol{X}$ be a random vector from an arbitrary n–dimensional distribution $\mathbf{F}$ and let F_i be the marginal distribution of X_i. Then, by Sklar (1959) there exists a function

$$C : [0,1]^n \to [0,1],$$

called a *copula* such that

$$F(x_1, \ldots, x_n) = C\big(F_1(x_1), \ldots, F_n(x_n)\big).$$

If F_i are continuous for every i, then the copula is unique and

$$C(u_1, \ldots, u_n) = F\big(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)\big),$$

where

$$F_i^{-1}(u) = \inf\left\{x \geq 0 | F_i(x) \geq 0\right\}.$$

For a copula the following properties hold:

- $C(u_1, \ldots, u_n) = 0$ whenever $u_i = 0$ for some i.
- $C(u_1, \ldots, u_n) = u_i$ whenever $u_j = 1$ for all $j \neq i$.
- $\max\left(\sum_{i=1}^n u_i + 1 - n, 0\right) \leq C(u_1, \ldots, u_n) \leq \min_{i=1,\ldots,n} u_i$, where the boundaries are called the *minimal copula* and the *maximal copula*.

Probably the most commonly used copula is *Gaussian copula* given by

$$C(u_1, \ldots, u_n) = \Phi_{\mathbf{R}}\big(\Phi_1^{-1}(u_1), \ldots, \Phi_n^{-1}(u_n)\big),$$

where Φ is the standard normal distribution function and $\Phi_{\mathbf{R}}$ is the cumulative distribution function of an n-dimensional normal distribution with zero mean and covariance matrix $\mathbf{R}$.

For more examples and theory about copulas, see Nelsen (2006).

C Frequency Data

Frequency data from Standard&Poor's (2006).

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	77	10	0	0	0	0	0	0
AA	4	189	17	0	0	0	0	0
A	0	22	434	32	1	0	0	0
BBB	0	0	13	244	14	0	0	0
BB	0	0	2	12	134	66	1	0
B	0	0	1	0	4	72	2	2
CCC	0	0	0	0	0	1	10	0

Table 1. S&P frequency matrix for year 1981.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	80	4	2	0	0	0	0	0
AA	1	204	15	1	3	0	0	0
A	0	18	409	45	3	0	0	1
BBB	1	0	6	243	26	1	0	1
BB	0	1	0	6	123	13	0	7
B	0	0	1	1	4	121	6	5
CCC	0	0	0	0	0	1	8	3

Table 2. S&P frequency matrix for year 1982.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	78	17	1	0	0	0	0	0
AA	1	228	13	2	0	0	0	0
A	3	19	404	20	3	0	0	0
BBB	1	2	19	249	18	2	0	1
BB	0	1	2	6	125	20	0	2
B	0	0	1	1	5	121	1	7
CCC	0	0	0	0	0	3	11	1

Table 3. S&P frequency matrix for year 1983.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	79	29	1	0	0	0	0	0
AA	5	271	13	3	0	0	0	0
A	0	13	426	16	3	0	0	0
BBB	0	1	35	232	18	6	0	2
BB	0	0	2	14	143	9	0	2
B	0	0	0	2	9	150	0	6
CCC	0	0	0	0	0	0	11	4

Table 4. S&P frequency matrix for year 1984.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	85	6	0	0	1	0	0	0
AA	1	282	27	6	0	4	1	0
A	0	10	440	34	6	1	0	0
BBB	0	2	23	216	18	12	0	0
BB	0	0	2	11	153	21	3	3
B	0	0	3	0	5	167	1	13
CCC	0	0	0	0	0	4	7	2

Table 5. S&P frequency matrix for year 1985.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	99	8	0	0	0	0	0	0
AA	4	311	16	6	0	2	0	0
A	1	26	436	50	8	9	0	1
BBB	0	0	21	224	25	8	1	1
BB	0	0	0	15	174	14	3	3
B	0	0	0	1	11	197	30	24
CCC	0	0	0	0	0	0	10	3

Table 6. S&P frequency matrix for year 1986.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	122	4	0	1	0	0	0	0
AA	6	318	18	1	0	0	0	0
A	0	7	429	30	2	6	0	0
BBB	0	2	17	249	21	11	0	0
BB	0	0	1	19	192	21	0	1
B	0	0	3	0	17	267	9	11
CCC	0	0	0	1	1	4	35	7

Table 7. S&P frequency matrix for year 1987.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	129	8	3	0	1	0	0	0
AA	5	297	38	10	2	1	0	0
A	0	8	444	23	4	3	0	0
BBB	0	1	29	244	17	7	2	0
BB	0	0	3	21	206	22	5	3
B	0	1	0	1	19	304	12	16
CCC	0	0	0	2	2	5	27	11

Table 8. S&P frequency matrix for year 1988.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	144	10	0	0	0	0	0	0
AA	2	324	24	0	0	0	0	0
A	0	8	474	36	12	1	0	0
BBB	0	0	23	260	20	2	2	2
BB	0	0	2	35	187	15	2	2
B	0	1	0	0	30	289	16	14
CCC	0	0	1	0	1	0	24	18

Table 9. S&P frequency matrix for year 1989.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	146	7	0	0	0	0	0	0
AA	2	336	39	0	0	0	0	0
A	0	12	478	42	7	1	0	0
BBB	0	0	17	292	18	3	0	2
BB	0	0	0	18	184	28	9	10
B	0	2	0	2	12	240	18	31
CCC	0	0	0	0	1	2	27	15

Table 10. S&P frequency matrix for year 1990.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	146	20	1	0	0	0	0	0
AA	1	356	30	0	0	0	0	0
A	1	3	519	35	3	0	0	0
BBB	0	3	14	313	19	3	0	3
BB	0	0	0	14	182	18	4	4
B	0	1	0	1	16	197	10	40
CCC	0	0	0	1	2	4	28	21

Table 11. S&P frequency matrix for year 1991.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	156	17	0	0	0	0	0	0
AA	5	408	29	5	0	0	0	0
A	0	7	585	26	1	1	0	0
BBB	0	1	18	341	15	3	0	0
BB	0	0	0	29	181	10	7	0
B	0	0	1	3	25	150	9	16
CCC	0	0	0	0	2	6	24	14

Table 12. S&P frequency matrix for year 1992.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	160	5	3	0	0	0	0	0
AA	0	440	28	1	0	0	0	0
A	2	6	612	25	0	0	0	0
BBB	0	0	20	379	27	0	1	0
BB	0	1	1	23	200	22	1	1
B	0	0	1	2	30	147	4	6
CCC	0	0	0	0	1	12	13	6

Table 13. S&P frequency matrix for year 1993.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	156	15	1	0	0	0	0	0
AA	2	434	43	0	0	0	1	0
A	0	9	681	30	1	1	1	1
BBB	0	1	14	446	10	2	0	0
BB	0	0	0	27	305	10	0	1
B	0	0	0	1	16	264	9	10
CCC	0	0	0	0	0	2	11	4

Table 14. S&P frequency matrix for year 1994.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	163	14	0	0	0	0	0	0
AA	3	436	46	1	0	0	0	0
A	0	20	793	31	2	0	0	0
BBB	0	2	25	518	21	0	0	1
BB	0	0	2	25	333	20	0	4
B	0	0	1	2	29	301	8	18
CCC	0	0	0	1	0	2	13	7

Table 15. S&P frequency matrix for year 1995.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	157	11	0	0	0	0	0	0
AA	3	445	26	0	0	0	0	0
A	0	28	859	19	1	0	0	0
BBB	1	0	40	613	12	1	0	0
BB	0	0	2	37	348	21	2	3
B	0	0	1	2	34	299	8	12
CCC	0	0	0	0	2	3	13	1

Table 16. S&P frequency matrix for year 1996.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	156	8	0	0	0	0	0	0
AA	4	469	19	4	0	1	0	0
A	0	18	901	42	2	3	0	0
BBB	0	1	28	693	19	6	0	2
BB	0	0	1	40	400	24	0	1
B	0	0	3	0	31	340	12	16
CCC	0	0	0	0	0	4	13	3

Table 17. S&P frequency matrix for year 1997.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	149	11	0	1	0	0	0	0
AA	2	465	27	0	0	0	0	0
A	1	15	897	59	3	0	0	0
BBB	0	0	24	798	50	7	2	4
BB	2	1	1	31	466	40	14	6
B	0	1	1	3	39	507	34	31
CCC	0	0	1	0	0	5	9	12

Table 18. S&P frequency matrix for year 1998.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	134	9	0	0	0	0	0	0
AA	1	479	38	3	0	0	0	1
A	0	24	898	61	1	1	0	1
BBB	0	2	30	855	40	0	0	2
BB	0	1	0	21	575	51	5	7
B	0	0	1	2	22	644	35	62
CCC	0	0	0	0	0	2	35	23

Table 19. S&P frequency matrix for year 1999.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	126	5	2	0	0	0	0	0
AA	4	465	60	1	0	0	0	0
A	0	25	890	79	4	1	0	1
BBB	0	3	27	934	38	5	3	4
BB	0	0	1	31	648	51	9	10
B	0	0	2	2	31	670	38	67
CCC	0	0	0	0	1	5	43	27

Table 20. S&P frequency matrix for year 2000.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	120	7	0	0	0	0	0	0
AA	1	447	59	0	0	0	0	0
A	0	25	916	74	2	0	4	2
BBB	0	1	37	990	52	7	11	4
BB	0	0	3	23	608	81	18	26
B	0	0	0	0	26	581	72	98
CCC	0	0	0	0	0	8	44	49

Table 21. S&P frequency matrix for year 2001.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	107	15	0	1	0	0	0	0
AA	1	394	89	12	1	3	0	0
A	0	6	931	118	9	2	1	1
BBB	0	1	27	1043	77	28	7	13
BB	1	0	3	26	623	61	11	22
B	0	0	0	2	35	515	69	61
CCC	0	0	1	0	2	13	58	75

Table 22. S&P frequency matrix for year 2002.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	92	9	2	0	0	0	0	0
AA	2	362	47	2	0	0	0	0
A	0	7	998	74	2	0	0	0
BBB	0	0	19	1158	68	4	0	3
BB	0	0	0	27	694	90	6	5
B	0	0	0	1	50	548	29	31
CCC	0	0	0	0	1	19	75	54

Table 23. S&P frequency matrix for year 2003.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	88	6	0	0	0	0	0	0
AA	1	368	15	1	0	0	0	0
A	0	13	1071	36	1	0	0	0
BBB	0	1	29	1311	27	2	0	0
BB	1	0	1	39	786	52	2	4
B	0	0	1	1	56	647	19	13
CCC	0	0	1	0	1	21	73	20

Table 24. S&P frequency matrix for year 2004.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	87	9	1	0	0	0	0	0
AA	0	369	20	2	0	0	0	0
A	1	20	1088	54	0	0	0	0
BBB	0	3	91	1290	47	7	0	1
BB	0	0	0	58	779	70	2	2
B	0	0	1	6	86	713	38	16
CCC	0	0	0	1	1	32	59	11

Table 25. S&P frequency matrix for year 2005.

Bibliography

Baxter, M. and A. Rennie (1996): *Financial Calculus: An Introduction to Derivative Pricing*. Cambridge University Press, Cambridge.

Bertoin, J. (1996): *Lévy Processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge.

Bertoin, J. (1999): Subordinators: examples and applications. In P. Bernard, editor, *Lectures on Probability Theory and Statistics: Ecole d'Ete de Probabilites de Saint–Flour XXVII — 1997*, volume 1717 of *Lecture Notes in Mathematics*, 4–91. Springer, Berlin.

Bielecki, T. R. and M. Rutkowski (2002): *Credit Risk: Modeling, Valuation and Hedging*. Springer Finance. Springer, Heidelberg.

Black, F. and J. C. Cox (1976): Valuing corporate securities: Some effects of bond indenture provisions. *The Journal of Finance*, **31**(2), 351–367.

Black, F. and M. S. Scholes (1973): The pricing of options and corporate liabilities. *Journal of Political Economy*, **8**(3), 637–654.

Bladt, M. and M. Sørensen (2005): Statistical inference for discretely observed Markov jump processes. *Journal of the Royal Statistical Society Series B*, **67**(3), 395–410.

Bluhm, C. and L. Overbeck (2007): Calibration of PD term structures: to be Markov or not to be. *Risk Magazine*, **20**(11), 98–103.

Bluhm, C., L. Overbeck and C. Wagner (2002): *An Introduction to Credit Risk Modeling*. Chapman & Hall/CRC, London.

Brent, R. P. (1973): *Algorithms for Minimization Without Derivatives*. Prentice–Hall.

Brigo, D. and F. Mercurio (2006): *Interest Rate Models — Theory and Practice: With Smile, Inflation and Credit*. Springer Finance. Springer, Berlin, 2nd edition.

Buhlmann, H., F. Delbaen, P. Embrechts and A. N. Shiryaev (1996): No arbitrage, change of measure and conditional Esscher transforms. *CWI Quarterly*, **9**(4), 291–317.

Carr, P. and L. Wu (2004): Time–changed Lévy processes and option pricing. *Journal of Financial Economics*, **71**(1), 113–141.

Carty, L. V. and J. S. Fons (1993): Measuring changes in corporate credit quality. Moody's Special Report, New York.

Çetin, U., R. Jarrow, P. Protter and Y. Yildirim (2004): Modeling credit risk with partial infromation. *The Annals of Applied Probability*, **14**(3), 1167–1178.

Chung, K. L. (2001): *A Course in Probability Theory*. Academic Press, San Diego, 3rd edition.

Collin-Dufresne, P., R. S. Goldstein and J. Helwege (2003): Is credit event risk priced? Modeling contagion via the updating of beliefs. Working paper.

Collin-Dufresne, P., R. S. Goldstein and J. N. Hugonnier (2004): A general formula for valuing defaultable securities. *Econometrica*, **72**(5), 1377–1407.

Cont, R. and P. Tankov (2003): *Financial Modelling with Jump Processes*. Financial Mathematics Series. Chapman & Hall/CRC Press, London, 1st edition.

Cox, J. C., J. E. Ingersoll and S. A. Ross (1985): A theory of the term structure of interest rates. *Econometrica*, **53**(2), 385–407.

Delbaen, F. and W. Schachermayer (2006): *The Mathematics of Arbitrage*. Springer Finance. Springer, Berlin.

Devroye, L. (1986): *Non–Uniform Random Variate Generation*. Springer.

Duffie, D., D. Filipovic and W. Schachermayer (2003): Affine processes and applications in finance. *Annals of Applied Probability*, **13**(3), 984–1053.

Duffie, D. and D. Lando (2001): Term structures of credit spreads with incomplete accounting information. *Econometrica*, **69**(3), 633–664.

Duffie, D. and K. J. Singleton (1999): Modeling term structures of defaultable bonds. *Review of Financial Studies*, **12**(4), 687–720.

Duffie, D. and K. J. Singleton (2003): *Credit Risk: Pricing, Measurement, and Management*. Princeton Series in Finance. Princeton University Press, Princeton (NJ).

Dupač, V. and J. Dupačová (1980): *Markovovy procesy II*. Universita Karlova, Praha.

Economic Research Division of Federal Reserve Bank of St. Louis (2008): Federal reserve economic data.
URL `http://research.stlouisfed.org/fred2/`

Elliott, R. J., L. Chan and T. K. Siu (2005): Option pricing and Esscher transform under regime switching. *Annals of Finance*, **1**(4), 423–432.

Federal Reserve Bank of New York (2008): Historical changes of the target federal funds and discount rates.
URL `http://www.ny.frb.org/markets/statistics/dlyrates/fedrate.html`

Feller, W. (1968): *An Introduction to Probability Theory and Its Applications*, volume I. John Wiley & Sons, New York, 3rd edition.

Filipovič, D. (2005): Time-inhomogeneous affine processes. *Stochastic Processes and their Applications*, **115**(4), 639–659.

Gerber, H. U. and E. S. W. Shiu (1994): Option pricing by Esscher transforms. *Transactions of Society of Actuaries*, **46**, 99–191.

Giesecke, K. (2004): Credit risk modeling and valuation: An introduction. In D. Shimko, editor, *Credit Risk: Models and Management*, chapter 16, 487–526. Risk Books, London, 2nd edition.

Giesecke, K. (2006): Default and information. *Journal of Economic Dynamics & Control*, **30**(11), 2281–2303.

Giesecke, K. (2008): Portfolio credit risk: Top down vs bottom up approaches. In R. Cont, editor, *Frontiers in Quantitative Finance: Credit Risk and Volatility Modeling*, Wiley Finance, chapter 10, 236–252. Wiley, Hoboken (NJ).

Grandell, J. (1976): *Doubly Stochastic Poisson Processes*, volume 529 of *Lecture Notes in Mathematics*. Springer, New York.

Guo, X., R. A. Jarrow and Y. Zeng (2008): Credit risk models with incomplete information. Working paper.

Hull, J. C. (2005): *Options, Futures and Other Derivatives*. Prentice Hall, New Jersey, 6th edition.

Inamura, Y. (2006): Estimating continuous time transition matrices from discretely observed data. Working paper, Bank of Japan.
URL `http://www.boj.or.jp/en/type/ronbun/ron/wps/data/wp06e07.pdf`

Israel, R. B., J. S. Rosenthal and J. Z. Wei (2001): Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. *Mathematical Finance*, **11**(2), 245–265.

Jacod, J. and A. Shiryaev (2003): *Limit Theorems for Stochastic Processes*, volume 288 of *A Series of Comprehensive Studies in Mathematics*. Springer, Berlin, 2nd edition.

Janssen, J. and R. Manca (2006): *Applied Semi–Markov Processes*. Springer, New York.

Jarrow, R. A., D. Lando and S. M. Turnbull (1997): A Markov model for the term structure of credit risk spreads. *Review of Financial Studies*, **10**(2), 481–523.

Jarrow, R. A. and S. M. Turnbull (1995): Pricing derivatives on financial securities subject to credit risk. *Journal of Finance*, **56**(5), 555–576.

Kallenberg, O. (1997): *Foundations of Modern Probability*. Springer, New York.

Karatzas, I. and S. E. Shreve (1991): *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer, New York, 2nd edition.

Karlin, S. and H. M. Taylor (1975): *A First Course in Stochastic Processes*. Academic Press, London, 2nd edition.

Karlin, S. and H. M. Taylor (1981): *A Second Course in Stochastic Processes*. Academic Press, London.

Kavvathas, D. (2001): Estimating credit rating transition probabilities for corporate bonds. Working paper, University of Chicago.

Kijima, M. (1997): *Markov Processes for Stochastic Modeling*. Stochastic Modeling Series. Chapman & Hall/CRC Press, London, 1st edition.

Kluge, W. (2005): *Time–inhomogeneous Lévy processes in interest rate and credit risk models*. Ph.D. thesis, University Freiburg.

Kreinin, A. and M. Sidelnikova (2001): Regularization algorithms for transition matrices. *Algo Research Quarterly*.

Kusuoka, S. (1999): A remark on default risk models. *Advances in Mathematical Economics*, **1**, 69–82.

Lando, D. (1998): On Cox processes and credit risky securities. *Review of Derivatives Research*, **2**(2–3), 99–120.

Lando, D. (2004): *Credit Risk Modeling: Theory and Applications*. Princeton Series in Finance. Princeton University Press, Princeton (NJ).

Lando, D. and T. M. Skødeberg (2002): Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking & Finance*, **26**, 423–444.

Lehmann, E. and J. P. Romano (2005): *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, New York, 3rd edition.

Lucas, A., A. Monteiro and G. V. Smirnov (2006): Nonparametric estimation for non–homogeneous semi–Markov processes: an application to credit risk. *Tinbergen Institute Discussion Papers*, (06-024/2).

McLachlan, G. J. and T. Krishnan (1997): *The EM Algorithm and Extensions*. Wiley, New York, 1st edition.

Merton, R. C. (1974): On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, **29**(2), 449–70.

Miyahara, Y. (2001): [Geometric Lévy process & MEMM] pricing model and related estimation problems. *Asia-Pacific Financial Markets*, **8**(1), 45–60.

Musiela, M. and M. Rutkowski (2005): *Martingale Methods in Financial Modelling*. Stochastic Modelling and Applied Probability. Springer, Berlin, 2nd edition.

Nelsen, R. B. (2006): *An Introduction to Copulas*. Springer Series in Statistics. Springer, New York, 2nd edition.

Prášková, Z. and P. Lachout (2005): *Základy Náhodných Procesů*. Karolinum, Praha.

Resnick, S. I. (1992): *Adventures in Stochastic Processes*. Birkhäuser, Boston, 1st edition.

Sato, K. (1999): *Lévy Processes and Infinitely Divisible Distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Schönbucher, P. J. (2003): *Credit Derivatives Pricing Models: Model, Pricing and Implementation*. Wiley, Hoboken (NJ).

Schoutens, W. (2003): *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley Series in Probability and Statistics. Wiley, London.

Schoutens, W. (2006): Exotic options under Lévy models: An overview. *Journal of Computational and Applied Mathematics*, **189**(1–2), 526–538.

Schuermann, T. (2004): What do we know about Loss Given Default? In D. Shimko, editor, *Credit Risk: Models and Management*, chapter 9, 249–274. Risk Books, London, 2nd edition.

Shiryaev, A. N. (1995): *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer, New York, 2nd edition.

Shreve, S. E. (2004): *Stochastic Calculus for Finance II: Continuous–Time Models*. Springer Finance. Springer, New York, 1st edition.

Sklar, A. (1959): Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229–231.

Standard&Poor's (2006): Annual 2005 global corporate default study and rating transitions. Technical report, Standard&Poor's, New York.
URL `http://www.domainb.com/finance/rating/standard_poor/S&P.doc`

Stroock, D. W. (2005): *An Introduction to Markov Processes*, volume 230 of *Graduate Texts in Mathematics*. Springer, Berlin.

Tuenter, H. J. H. (2001): The minimum L_2-distance projection onto the canonical simplex: A simple algorithm. *Algo Research Quarterly*, **4**(4), 53–55.

Vašíček, O. A. (1977): An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**(2), 177–188.

Vašíček, O. A. (1984): Credit valuation. Working Paper, KMV Corporation.

Van Loan, C. (1978): Computing integrals involving the matrix exponential. *IEEE Transactions on Automatic Control*, **23**(3), 395–404.

Zhou, C. (1997): A jump–diffusion approach to modeling credit risk and valuing defaultable securities. Finance and Economics Discussion Series 1997–15, Board of Governors of the Federal Reserve System.